Semantics and the electrophysiology of meaning

Tense, aspect, event structure

Giosuè Baggio

ISBN 978-90-9023969-9

Printed by Ipskamp Drukkers

Semantics and the electrophysiology of meaning

Tense, aspect, event structure

Een wetenschappelijke proeve op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann
volgens besluit van het college van decanen
in het openbaar te verdedigen op 6 mei 2009
om 15:30 uur precies

door

Giosuè Baggio

geboren op 6 april 1979
te Palmanova

Promotors:
Prof. dr. P. Hagoort
Prof. dr. M. van Lambalgen (University of Amsterdam)

Manuscript committee:
Prof. dr. H. Schriefers
Prof. dr. K. Stenning (University of Edinburgh)
Dr. J. van Berkum (Max Planck Institute for Psycholinguistics)

Contents

Foreword

We are capable of producing and comprehending utterances that are novel not only in their phonologic and syntactic forms but also, perhaps most characteristically, in their meanings. This capacity is largely unique to humans, appears at a particular stage of development, and may deteriorate following traumata or pathologies that often have neurological origins. It is therefore a platitude that a mature, normal-functioning brain is necessary for us to possess meanings and do with them whatever we do in ordinary discourse. What is far less obvious is how the brain does it, and that is precisely the issue here. This book approaches linguistic meaning in the human brain in a principled manner, combining modeling tools from formal logic with experimental techniques from neurophysiology, and using the semantics of tense, aspect and event structure as a test case.

What principles underlie our ability to code/decode meanings to/from speech signals? Different answers have been given within different areas of cognitive science. Experimental work in psychology and neuroscience has found support for such notions as *immediacy* and *incrementality*, which appear descriptively adequate but resist rigorous formalization. Logic and linguistics, on the other hand, have focused on desiderata for theories of meaning such as *compositionality* and *monotonicity*, which afford precise formal analyses but provide little empirical coverage as far as language comprehension, as opposed to entailment relations, is concerned. This seeming trade-off between formal and empirical adequacy makes a shared framework for the characterization and testing of these principles appear beyond reach. Yet, it seems reasonable to seek some form of unification of research programs dealing with the same questions, albeit from different perspectives. For that purpose, one needs a pluralistic philosophy of science – presented in chapter 1 – which is more concerned with establishing connections between theories, and making adjustments in each, than with trying to reduce one theory to another, or even to eliminate one in favor of the other. Logic and neuroscience are developing accounts of meaning that are legitimate in their own right, and furthermore their relations fall outside all available schemes for reduction. So, talking of inter-theoretic bridging instead of reduction makes the task of designing a unified framework for the study of meaning less daunting. Chapters 2-6 present some modeling and experimental work that exemplifies a viable integrative methodology.

The first part of the book, 'from semantics to neuroscience', discusses some ins and outs of bringing formal semantics – and the principles of compositionality and monotonicity in particular – to bear on processing data. The motivation for the whole enterprise is presented in chapter 1, which also anticipates most if not all the empirical issues discussed in the rest of the book: time and causation in language, the 'binding problem for semantics' etc. That chapter however is chiefly a philosophical and methodological one, in that it pins down and discusses choices that have led to the current state of separation of disciplines investigating meaning: the competence/performance distinction, the use of intuitions in linguistics and psycholinguistics etc. We argue that formal semantics can constrain predictions in language processing research, and that processing data can be used to correct semantic theories. This two-way route from semantics to neuroscience is usually seen as a dead end, and charges of 'psychologism', 'categorial errors' and the like are sometimes issued. However, David Marr has shown how optics and solid geometry can be included in a theory of visual processing, and much in the same way logic can be part of a theory of meaning processing. Thus, if one takes formal semantics as a collection of 'computational theories' in Marr's sense, one can ask what their algorithmic and physical implementations look like, and therefore how they play out in processing reality. As we argued above, this has little to do with psychologism, eliminativism, and reductionism. Rather, it is an attempt at drawing connections between whatever approximates – to some reasonable degree of formalization – a theory of meaning at each level of analysis in Marr's scheme. Such a connection-drawing exercise is carried out for monotonicity and compositionality in chapters 2 and 3 respectively. Chapter 2 is more limited in its scope, as it discusses the processing consequences of monotonicity focusing on one particular linguistic phenomenon: the imperfective paradox. However, more examples of non-monotonic discourse processing are provided in chapter 1. Chapter 3 examines the processing consequences of compositionality using a somewhat broader array of linguistic and psycholinguistics examples, while keeping tense, aspect and event structure under the spotlight. In chapters 1-3, immediacy and incrementality are assumed as further constraints on the scope of simple composition and monotonic inference. Moreover, they are integrated into discourse semantics in a more radical way than has been previously done, for instance with the notion of incrementality in DRT.

The second part, 'the electrophysiology of meaning', reports three ERP experiments on discourse comprehension in which predictions and explanations are inspired by the formal analyses of part 1. These studies are the benchmark test for both the methodology of chapter 1 and the account of meaning in the brain of chapters 1-3. What does the latter look like? We see discourse processing as highly incremental, opportunistic as it uses first whatever type of information is available first (immediacy), non-monotonic as it allows the recomputation of discourse models, and non-compositional in that the sources of input to on-line meaning composition are not restricted to syntax and the lexicon. Each of the ERP studies reported in the second part was designed to validate a bit of this view. The first experiment, on tense violations (chapter 4), presents some ERP evidence for immediacy, and accounts for that in a constraint-based theory of early stages of semantic processing. The second is an experiment on aspect processing (chapter 5) in which ERP data are shown that are consistent with the non-monotonic theory of the imperfective paradox described in chapter 2. The third study, on complement coercion (chapter 6), presents evidence for enriched semantic composition in computing event structures. The conclusions look back at the ERP findings in the light of considerations of processing architecture, add some neural detail to the picture, and lay down a tentative path for future work.

Acknowledgments

I wish to thank Peter Hagoort and Michiel van Lambalgen for helping me turn the privilege of graduate studies into a meaningful and valuable experience. They tailored this doctorate to my preferences and skills, and gave me a large amount of freedom in whatever required my contribution. To anyone knowing the work of Michiel and Peter, it should be clear how much the studies presented here are indebted to them. Several other people supported in many ways the production of this dissertation. I want to mention them here, and emphasize how much each chapter has improved owing to their suggestions and criticism: Oliver Bott, Emmanuel Chemla, Travis Choma, Bart Geurts, Friederich Hamm, Wilfrid Hodges, Peter Indefrey, Miriam Kos, Laura Menenti, David Neville, Ira Noveck, Karl Magnus Petersson, Judith Pijnacker, Liina Pylkkänen, Brett Reynolds, Tineke Snijders, Mark Steedman, Keith Stenning, Martin Stokhof, Jos van Berkum, Tessa van Leeuwen, Theo Vosse, Roel Willems, Henk Zeevat, and several anonymous reviewers. Herbert Schriefers, Keith Stenning and Jos van Berkum graciously served as a thesis committee. I am grateful to NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) for financing the research reported here.

Part 1

From semantics to neuroscience

CHAPTER 1

Philosophy and methodology

This chapter is a modified version of G. Baggio, M. van Lambalgen & P. Hagoort. Language, linguistics and cognition. In R. Kempson & T. Fernando (Eds.), *Handbook of Philosophy of Linguistics*. Volume 14 of D. Gabbay, P. Thagard & J.H. Woods (Eds.), *Handbook of Philosophy of Science*. Elsevier, to appear.

1. Introduction

Experimental research during the last few decades has provided evidence that language is embedded in a mosaic of cognitive functions. An account of how language interfaces with memory, perception, action and control is no longer beyond the scope of linguistics, and can now be seen as part of an explanation of linguistic structure itself. However, although our view of language has changed, linguistic methodology is lagging behind. This chapter is a sustained argument for a diversification of the kinds of evidence applicable to linguistic questions at different levels of theory, and a defense of the role of linguistics in experimental cognitive science.

1.1. Linguistic methodology and cognitive science. At least two conceptual issues are raised by current interactions between linguistics and cognitive science. One is whether the structures and rules described by linguists are cognitively real. There exist several opinions in this regard, that occupy different positions on the mentalism/anti-mentalism spectrum. At one extreme is cognitive linguistics [**31**], endorsing both theoretical and methodological mentalism. The former is the idea that linguistic forms are related causally and conceptually to other mental entities. The latter calls for a revision of traditional linguistic methodology, and emphasizes the role of cognitive data in linguistics. At the opposite side of the spectrum lies formal semantics which, inspired by Frege's anti-psychologistic stance on meaning and thought [**60, 125, 15**], rejects both versions of mentalism. Somewhere between the two poles is Chomsky's [**23**] theoretical mentalism, which sees linguistic rules as ultimately residing in the brain of speakers. However, his commitment to the cognitive reality of grammar does not imply a revision of linguistic methodology, which is maintained in its traditional form based on native speakers' intuitions and the competence/performance distinction.

The second problem – partly dependent on the first – is whether experimental data on language acquisition, comprehension and production have any bearing on linguistic theory. On this point too, there is no consensus among linguists. The division between competence and performance has often been used to secure linguistics from experimental evidence of various sorts [**14**], while intuitive judgments of native speakers were regarded as the only type of data relevant for the theory [**23**]. However, some authors have granted that at least behavioral data should be allowed to inform competence theories, for instance if the linguist is studying a language which is not her own native language [**128**]. Others have proposed frameworks in which competence can be connected to performance mechanisms [**99**]. But while these models account for how competence constrains performance [**101**], they seem to overlook the possibility that the reverse is also the case: aspects of linguistic structure may be the outcome of evolutionary processes leading to an adaptation of the brain to language use, that is to performance [**164**]. Generative grammar and formal semantics have regarded accounts of competence as shielded from data provided by experimental psychology and neuroscience. A more inclusive attitude has been adopted by psycholinguists and cognitive brain scientists, driven by an increasing demand of theories and models that would account for their data [**16, 46, 64, 139, 140**]. But despite these attempts, a methodological framework relating linguistics, language processing and low-level neural models is still missing.

1.2. Language, lower and higher cognition. Most theories in cognitive linguistics and neuroscience regard language as grounded in perception and action. For instance, cognitive semanticists have proposed that the meanings of concrete nouns stored in memory include stereotyped visual-geometric representations of the entities they refer to [**97**]. Analogously, representations of action verbs might embrace aspects of the relevant motor programs [**70**]. It has also been suggested that the building blocks of semantics like the predicate-argument structure originate in the functional and anatomical organization of the visual and auditory systems [**94**]. Experimental work in cognitive neuroscience indicates that language has ties with the sensory-motor systems, but methodology, specific data points and accounts of how exactly language connects to 'lower' cognition are still debated [**168, 167, 169, 51, 86, 95, 156, 200, 199**]. The interested reader may want to follow further these references: in this chapter we will focus on language and 'higher' cognitive domains such as planning and reasoning. A motivation for this choice is that planning and reasoning shed light on the computation of *complex* linguistic structures, which is where language really comes into its own, whereas looking at the interactions between language and the sensory-motor systems gives us more insights into representations and processes at the *lexical* level.

It has been proposed that the recursive organization of plans supplies a mechanism for combinatory operations in grammar [**191**], and the goal-directed nature of planned actions constrains cognitive constructions of time, causality and events, with consequences for the semantics of tense, aspect and modality [**191, 212**]. Planning might as well be implicated in the production and comprehension of discourses. Language processing requires adjusting the current discourse model incrementally given the input. If further information counters earlier commitments or expectations, a recomputation of the initial discourse model may be necessary to avoid inconsistencies. This process is best accounted for by the non-monotonic logic underlying planning, and more generally executive function, where the chosen action sequence may be readjusted when obstacles are encountered along the way.

On a par with planning, reasoning is of special interest in this chapter. Some have seen non domain-specific thought and reasoning as the most notable among the cognitive skills subserved by language [**18, 19**]. This suggestion is sometimes implicit in logical approaches to language since Boole [**9**, Ch. 2, p. 24] and bears some resemblance to the psycholinguistic notion that reasoning follows and builds upon interpretation [**105, 187**]. In this perspective, interpretation equals reading off of logical (often classical) form from a sentence's surface structure for subsequent elaboration involving inference. So there is a one-way dependency of reasoning from interpretation: interpretation supports reasoning, though not vice versa. Others have seen the relation between interpretation and inference as based on a two-way interaction [**194**]: reasoning is involved in computing a model of what is said *and* in deriving conclusions from it. Human communication is thus regarded as the foremost skill enabled by language, and reasoning serves the purpose of coordinating different interpretations of an utterance or different situation models across speakers [**193**].

2. Linguistics and cognitive data

2.1. A methodological paradox. It is often said that the relations between cognitive science and linguistics began to be fully appreciated only after the publication of Chomsky's early writings, and in particular *Aspects of the Theory of Syntax* in 1965. This is certainly true if what is at stake is theoretical mentalism – the notion that linguistic theory deals ultimately with a system of representations and rules in the speaker's mind/brain. However, although this particular form of theoretical mentalism encourages and to some extent requires some interaction between the two disciplines, the choice of regarding the study of competence as in principle indifferent to the results of experimental research had the opposite effect, that of separating theories of meaning and grammar from models of language processing. Many would agree that the contacts between linguistics and cognitive psychology have not been as deep and systematic as they could have been, had various obstacles to fruitful interaction been removed. What is more difficult to appreciate is the existence of a tension in the very foundation of generative grammar, and the inhibiting effect it had on the growth of linguistics within cognitive science. Before we move on, it may be worth recovering the terms of this 'paradox' directly from Chomsky's text.[1]

One side of the dilemma is represented by a number of remarks contained in §1 of the first chapter of *Aspects*, where Chomsky writes:

> We thus must make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations). Only under [. . .] idealization [. . .] is performance a direct reflection of competence. In actual fact, it obviously could not directly reflect competence. A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on. The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior. [**23**, p. 4]

The task of the linguist is that of providing an account of competence based on performance data, that is on normalized records of linguistic behavior. Chomsky grants that performance data are essential to linguistic theorizing. But the issue to be settled, which in fact lies at the heart of the paradox, is exactly what counts as linguistic behavior, or more precisely what kind of performance data can form the empirical basis of competence theories. Generative linguists would contend that it was never a tenet of their research program to admit data other than native speakers' intuitions, but this is not what Chomsky's remarks suggest. On the contrary, he seems to admit a variety of data types:

> Mentalistic linguistics is simply theoretical linguistics that uses performance as data (along with other data, for example, the data provided by introspection) for the determination of competence, the latter being taken as the primary object of its investigations. [**23**, p. 193]

The evidential base of linguistics consists of introspective judgments *and* performance data, that Chomsky mentions here as if they were in an important sense different from intuitions. Moreover, intuitions are alluded to here as a subsidiary source of evidence, and as part of a larger class of data types. The question is precisely what should be considered performance data. Is elicited and experimentally controlled behavior allowed to exert some influence on accounts of competence? There are reasons to believe that Chomsky would have answered "yes", the most important of which has to do with his remarks on the limits of intuitions. In 1955, in *The Logical Structure of Linguistic Theory* [**21**] he wrote:

> If one of the basic undefined terms of linguistic theory is 'intuition', and if we define phonemes in this theory as elements which our intuition perceives in a language, then the notion of phoneme is as clear and precise as is 'intuition'. [...] It should be clear, then, why the linguist interested in constructing a general theory of linguistic structure, in justifying given grammars or (to put the matter in its more usual form) in constructing procedures of analysis should try to avoid such notions as 'intuition'. [**21**, pp. 86-87]

An even more explicit position was expressed in the 1957 book *Syntactic Structures*, where Chomsky suggests that hypotheses on properties of linguistic strings and their constituents should be evaluated on the basis of controlled operational tests. Relying on native speaker's judgments or intuitions, he writes,

---

[1]Over the years Chomsky has entertained different opinions on these issues. Here we choose to focus on those expressed in *Aspects of the Theory of Syntax* [**23**] because these have probably been the most influential. So we identify Chomsky with this particular text rather than with the actual linguist.

> amounts to asking the informant to do the linguist's work; it replaces an operational test of behavior (such as the pair test) by an informant's judgment about his behavior. The operational tests for linguistic notions may require the informant to respond, but not to express his opinion about his behavior, his judgment about synonymy, about phonemic distinctness, etc. The informant's opinions may be based on all sorts of irrelevant factors. This is an important distinction that must be carefully observed if the operational basis for grammar is not to be trivialized. [**22**, pp. 8-9][2]

Controlled operational tests are thus necessary in order to overcome the difficulties arising from relying exclusively on native speakers' intuitions. This implies that introspective data are dismissed as an inadequate source of evidence for linguistic theory. So here is one horn of the dilemma: mentalistic linguistics rejects speakers' intuitions and requires performance data, including controlled behavioral tests, to constrain the theory of competence.

The other side of the paradox is represented by a series of remarks in §4 of chapter 1 of *Aspects*, where Chomsky questions the nature of the empirical basis of competence theories:

> There is, first of all, the question of how one is to obtain information about the speaker-hearer's competence, about his knowledge of the language. Like most facts of interest and importance, this is neither presented for direct observation nor extractable from data by inductive procedures of any known sort. Clearly, the actual data of linguistic performance will provide much evidence for determining the correctness of hypotheses about underlying linguistic structure, along with introspective reports (by the native speaker, or the linguist who has learned the language). [**23**, p. 18]

Experimental research based on controlled observation and statistical inference is seen as providing facts of no 'interest and importance', and rejected as ineffective for the purposes of the theory of competence. Interestingly, intuitions are treated as they were on a par with performance data. Not for long, however, because Chomsky a few paragraphs later takes an important step away from psychology:

> The critical problem for grammatical theory today is not a paucity of evidence but rather the inadequacy of present theories of language to account for masses of evidence that are hardly open to serious question. The problem for the grammarian is to construct a description and, where possible, an explanation for the enormous mass of unquestionable data concerning the linguistic intuition of the native speaker (often, himself); the problem for one concerned with operational procedures is to develop tests that give the correct results and make the relevant distinctions. [...] We may hope that these efforts will converge, but they must obviously converge on the tacit knowledge of the native speaker if they are to be of any significance. [**23**, pp. 19-20]

The door that seemed to be open for controlled experimentation is now closed. The range of data that could affect the theory of competence has been narrowed down to intuitions, and more specifically to those of the linguist. The task of experimental research, Chomsky says, is to develop tests that would ultimately align with introspective data. The convergence of linguistics and psychology is thus projected forward in time as a desirable outcome not of the joining of efforts, but of their strict segregation. Not only are linguistics and psychology now regarded as separate enterprises, but psychology is also required – in order to meet a standard of explanatory adequacy – to provide results that are consistent with the theory of competence as based on the linguist's intuitions.

We should like to note that, whatever the methodological choices made by linguists, this seems an unacceptable requirement for psycholinguists and neuroscientists alike, who should instead be aided to deliver reliable data, no matter which theory or whose intuitions

[2]The circularity which Chomsky is alluding to here is also mentioned by Quine in his 1970 paper on linguistic methodology: "We are looking for a criterion of what to count as the real or proper grammar, as over against an extensionally equivalent counterfeit. [...] And now the test suggested is that we ask the native the very question we do not understand ourselves: the very question for which we ourselves are seeking a test. We are moving in an oddly warped circle." [**174**, p. 392].

are eventually supported. If experimental research provides data which do not align with the introspective judgments of linguists or other native speakers, then, following common practice in science, there seems to be no other choice than to accept empirical evidence and reject intuitions. What is perhaps most disturbing is Chomsky's disregard toward any form of experimental testing in linguistics:

> In any event, at a given stage of investigation, one whose concern is for insight and understanding (rather than for objectivity as a goal in itself) must ask whether or to what extent a wider range and more exact description of phenomena is relevant to solving the problems that he faces. In linguistics, it seems to me that sharpening of the data by more objective tests is a matter of small importance for the problems at hand. [**23**, p. 20]

The second horn of the dilemma is the following: linguistic theory is based primarily on the intuitions of native speakers, and does not require controlled experimentation to constrain accounts of competence.

2.2. The vagaries of intuition. For some linguists, in particular in generative grammar and formal semantics, the intuitions of native speakers constitute the empirical basis of the theory of competence. But the prominent place assigned to intuitions by modern linguistic methodology seems at odds with the principles of mentalism. If competence is a system of rules and structures realized in the speaker's brain, and if behavior reflects the functioning of such system, then a linguist constructing a competence theory – and perhaps analogously a child learning a language – must solve an 'inverse problem', that of inferring the rules of competence from observable performance. In order to solve this problem, the linguist might need to take into account a broad range of data. So *any* reliable physiological or behavioral measure of performance should, at least in principle, be allowed to contribute to the theory of competence. The question is where should one draw a line between relevant (intuitions?) and irrelevant (neurophysiology?) data, and why. Until convincing answers are found, the more comprehensive one's methodological framework, the better. Here is why mentalism is to be preferred over traditional philosophies of language.

The conflict with mentalism is however not the only problem raised by introspective judgments. Another source of concern is Chomsky's claim that intuitions are not only the starting point of linguistic theorizing, but also the standard to which any grammar should conform:

> A grammar can be regarded as a theory of language; it is *descriptively adequate* to the extent that it correctly describes the intrinsic competence of the idealized native speaker. The structural descriptions assigned to sentences by the grammar, the distinctions that it makes between well-formed and deviant, and so on, must, for descriptive adequacy, correspond to the linguistic intuition of the native speaker (whether or not he may be immediately aware of this) in a substantial and significant class of crucial cases. [**23**, p. 24]

Supposing the tension with mentalism were relieved, allowing other data types to influence competence models, and introspective judgments were used only at the outset of linguistic inquiry, intuitions would still pose a number of serious methodological problems. It is not just the role of intuitions in linguistic theorizing that has to be put under scrutiny, but also the very claim that intuitions offer a vantage point on tacit knowledge.

2.2.1. *Intuitions in linguistics.* If the system of linguistic rules in a speaker's brain really is "deeply unconscious and largely unavailable to introspection" [**100**, p. 652], one should see discrepancies between overt linguistic behavior, that reflects 'unconscious' competence rules, and the intuitions or beliefs that speakers have about these rules. This notion has been substantiated by Labov [**119**], who collected evidence on a wide variety of cases in regional American English. One example is the positive use of 'anymore' in various sections of the Philadelphia white community, meaning that a situation which was not true some time in the past is now true, roughly equivalent to 'nowadays':

(1) Do you know what's a lousy show anymore? Johnny Carson.

Labov interviewed twelve speakers who used the adverb freely and consistently with its vernacular meaning exemplified in (1). He reported a majority of negative responses when they were asked whether a sentence like (1) is acceptable, and surprisingly weak intuitions on what the expression signifies in their own dialect, which contexts are appropriate for its use, and what inferences can be drawn from its occurrences.

There are several other arguments for the unreliability of intuitions in linguistics. For instance, Marantz [**128**] has observed that grammaticality is a technical term defined within linguistic theories: a sound/meaning pair is grammatical or well-formed with respect to a grammar if and only if that grammar generates or assigns a structural description to the pair such that all relevant grammaticality or well-formedness constraints can be satisfied. In the quote from *Aspects* above, Chomsky takes for granted that structural descriptions assigned by some grammar to sentences can be checked for correspondence against native speakers' judgments. However, native speakers of a language can hardly be said to have intuitions of grammaticality in the technical sense, nor do they grasp other properties of strings as they are defined within a formal grammar. Moreover, naïve language users might conflate into the notion of grammaticality different morpho-syntactic, semantic and pragmatic criteria of felicitousness, and they might do so in a way that is beyond control for the linguist. Similar observations would also apply to intuitive judgments of synonymy or truth-conditions, as opposed to formal definitions within a semantic theory.

As a way out, one might argue that a caveat only applies to naïve informants, and that the intuitions of linguists, immune to pre-theoretical notions of grammaticality, synonymy, and the like, are in fact reliable [**38**]. Relevant to this issue, is an experiment by Levelt [**123**] in which the intuitions of twenty-four trained linguists were investigated. Participants were presented with fourteen examples from their own field's literature, among which:

(2) a. No American, who was wise, remained in the country.
b. The giving of the lecture by the man who arrived yesterday assisted us.

None of the linguists rated correctly the ungrammatical sentence (2a), and sixteen judged the well-formed sentence (2b) as ungrammatical. Ungrammatical sentences had less chance of being judged ungrammatical than grammatical items. Levelt warns against taking these results too seriously, but he observes with some reason that "they are sufficiently disturbing to caution against present day uses of intuition" [**123**, p. 25].

We could go on providing other examples of the problems that might arise with the use of introspective reports in the analysis of specific natural language sentences. However, we should now like to take a different approach, considering an argument targeted at the very nature and scope of intuitions. The argument, introduced and discussed by Hintikka [**90**], starts with the observation that intuitions of grammaticality, synonymy etc. always relate to particular sentences (i.e. tokens), and not to entire classes of items or the common syntactic or semantic structure they share (i.e. types). Hintikka writes that

> intuition, like sense perception, always deals with particular cases, however representative. [...] But if so, intuition alone cannot yield the general truths: for instance, general theories for which a scientist and a philosopher is presumably searching. Some kind of generalizing process will be needed, be it inductive inference, abduction, or a lucky guess. The intuitions [Chomsky] recommended linguists to start from were intuitions concerning the grammaticality of particular strings of symbols, not concerning general rules of grammar. [**90**, p. 137-138]

Against Hintikka's claim, one may argue that also paradigmatic variation is a proper object of intuition. The linguist would then be able to generalize over the properties of linguistic structures by constructing a paradigmatic set of sentences exhibiting those properties. This view however can be countered with the observation that the supposed 'intuitions' about paradigmatic cases are rather more similar to theory-laden hypotheses than to introspective judgments of naïve informants. The linguist, in order to construct such paradigmatic items, has to be able to control all irrelevant variables and systematically manipulate the factors of interest. This, in turn, requires that the linguist knows details of the grammar or the logical structure of the language which seem inaccessible to naïve speakers. It is this knowledge, which is often drawn from existing theories, that allows the linguist to have intuitions about linguistic structure. And this leads us to Hintikka's key statement, that competence theories are not equipped with built-in devices for deriving abstract grammatical or semantic forms from particular linguistic samples. That is,

> reliance on generalization from particular cases is foreign to the methodological spirit of modern science, which originated by looking for dependencies of different factors in instructive particular cases (often in controlled experimental situations), and by studying these dependences by the same mathematical means as a mathematician uses in studying the interdependencies of different ingredients of geometrical figures in analytic geometry. [...] transformational grammarians and other contemporary linguists would do a much better job if, instead of relying on our intuitions about isolated examples, they tried to vary systematically suitable ingredients in some sample sentence and observed how our 'intuitions' change as a consequence. Now we can see why such systematic variation is a way of persuading our intuitions to yield general truths (dependence relations) rather than particular cases. [**90**, p. 135]

If intuitions are to serve as a reliable starting point in linguistic inquiry, then they should be proved to have systematic properties. Observing patterns of covariation of introspective judgments and other factors – such as the structure, the content, the context of occurrence of the sentence, the attitude of the speaker, and so on – would make the particular examples under consideration instructive and thus effective as part of the empirical basis of linguistic theories. The important consequence is that, in order to systematically alter the ingredients of sample sentences, the linguist should be able to control these factors in a manner similar to the manipulation of experimental variables in laboratory research. The solution offered by Hintikka to the problem of intuitions points in the direction of *infusing linguistic practice with psychological experimentation*. The linguist would as usual start from intuitions, but only the systematic aspects of these as revealed by experimentation, and if necessary statistical tests, would be preserved and transferred into the theory (see [**14**, pp. 158-163] for a similar position).[3] Hintikka offers an intriguing example, in which one tries to define the meaning of an expression in Montague grammar on the basis of systematic dependencies between subjects' intuitions and the contexts of occurrence of the expression of interest. In particular, he writes, if the notion of possible world is allowed in the theory,

> then there is, in principle, no definite limit as to how far your experimentation (construction of ever new situations) can carry you in determining the class of scenarios in which the word does or does not apply. And such a determination will, at least for a Montagovian semanticist, determine the meaning of the word. Indeed, in Montague semantics, the meaning of a term is the function that maps possible worlds on references (extensions) of the appropriate logical type (category). And such functions can, in principle, be identified even more and more fully by systematic experimentation with the references that a person assigns to his terms in different actual or imagined scenarios. [**90**, p. 146][4]

However, it may be a fair point in favor of introspective judgments in a broader sense to add that Hintikka considers thought experiments on a par with genuine experimentation [**90**,

[3]Bunge [**14**, p. 168] pinpoints several methodological choices in generative linguistics which seem to diminish its relevance in empirical science, such as the "conduct of linguistic inquiry in total independence from neuroscience, social science, and even scientific psychology" and "a heavy reliance on intuition". We too consider these as obstacles to understanding language, but we disagree with the judgment that Bunge formulates based on these remarks – that modern linguistics is (or has been) pseudo-scientific.

[4]Although we consider Hintikka's an informative example of linguistic theorizing based on covariation patterns of contextual factors and intuitions, we must also add that there are serious problems with the notion of meaning (that is, Frege's *Sinn*) in Montague semantics. For instance, since the theory allows for infinitely many possible worlds, it becomes unclear whether we can even approximate the meaning of an expression using Hintikka's method.

p. 146]. Thus, instead of eliciting overt responses from subjects in a number of conditions, the experimenter imagines herself in such situations. If the relevant variables are controlled with as much care as one would exercise in an experimental setting, introspection can reveal systematic aspects of language use, and thus contribute to theories of competence.

Hintikka's argument can be made more explicit with reference to a number of studies investigating the role of the most important of his 'suitable ingredients' – context. Linguistic and psycholinguistic research has demonstrated that the context in which a sentence occurs can affect judgments of acceptability. Bolinger [**8**] reported that sentences, which speakers judge as semantically implausible when presented in isolation, are regarded as acceptable when embedded in context. Consider the following examples:

(3) a. It wasn't dark enough to see.
b. I'm the soup.

These sentences are typically judged as semantically deviant, although for different reasons: (3a) because one normally needs light in order to see, and (3b) because the predicate 'being a soup' cannot be applied to a human being. Now, consider the same sentences embedded in a suitable discourse context, with (4b) being spoken at a cashier's counter in a restaurant:

(4) a. I couldn't tell whether Venus was above the horizon. It wasn't dark enough to see.
b. You've got us confused: you're charging me for the noon special. The man in front of me was the noon special. I'm the soup.

Examples (3) in an appropriate context seem perfectly acceptable. Because context has such marked effects on intuitions, linguistic theory, if it has to rely on introspective judgments, should explicitly take into account this fact. *Mutatis mutandis*, a similar observation applies to experimental research on language comprehension and production.

2.2.2. *Intuitions in psycholinguistics.* The appeal to intuitions was not an explicit choice of methodology in psycholinguistics and the cognitive neuroscience of language. In fact, the method of introspection was discarded in scientific psychology after its failures in the 19th century. However, it is adopted in language processing research as a means of establishing differences between sentence types to be used as stimuli in actual experiments. The typical procedure for setting up a language processing study is to start with a few sample sentences differing with respect to some linguistic feature, the assessment of which is initially left to the intuitions of the experimenter. For instance, let us consider one of the first ERP studies on syntactic constraint violations, by Osterhout and Holcomb [**151**]. Here the starting point is a pair – or a relatively small set of pairs – of sentences containing either an intransitive (5a) or a transitive (5b) verb, using a direct object construction:

(5) a. The woman struggled to prepare the meal.
b. The woman persuaded to answer the door.

Up to this stage, the methodology is by and large the same as that of the linguist. However, while the latter would then proceed with, say, formalizing the requirements of intransitive and transitive verbs with respect to direct objects, the psycholinguist, to make sure there is sufficient statistical power to test her processing hypotheses in a dedicated experiment, would have to construct a large set of sentences with the same structure and properties of (5a-b). In the next step, the sentences would be presented to subjects while the dependent variables of interest are measured, which in the study of Osterhout and Holcomb were ERPs and grammaticality ratings. Grammatical sentences like (5a) were judged to be acceptable in 95% of the cases, and supposedly ungrammatical items like (5b) in 9% of the cases. One may argue, as an academic exercise towards an explanation of the 9% figure, that (5b) does have contexts in which it is both grammatical and semantically acceptable, for instance if it is interpreted as a reduced relative clause ('The woman who was persuaded to answer the door'), and is uttered as an answer to a *who* question, as in the following dialogue:

(6) A: Who stumbled on the carpet in the hallway?
B: The woman persuaded to answer the door.

We have already encountered this phenomenon discussing Bolinger's examples above. In a context such as (6), Osterhout and Holcomb's ungrammatical sentence becomes perfectly admissible. Acceptability judgments, therefore, depend on the range of uses (or contexts of use) readers are willing to consider. In this sense, subjects' intuitions may differ from those of the experimenter. For example, a linguist would remind us that 'The woman persuaded to answer the door' is an NP, and not a sentence. But what prevents naïve language users from including well-formed NPs into their notion of 'sentence'? Here the answer can only be the *linguist's* own notion of 'sentence'. This also suggests that discrepancies between the intuitions of naïve informants and trained scientists may be more important than isolated linguists' intuitions when it comes to fully explaining a data set.

2.3. Beyond competence and performance. Intuitions are but one of the many sources of concern for a thoroughly mentalistic approach to language. As Jackendoff [**99**, p. 29] has pointed out, there is a conflict, which roughly coincides with the dilemma as we described it above, between Chomsky's theoretical mentalism and traditional linguistic methodology as based on intuitions and on the competence/performance distinction. Mentalism requires at least a revision of that distinction. Jackendoff [**99**] has addressed this issue, trying to find a more accommodating formulation which allows a natural interplay of linguistics and the cognitive sciences:

> Chomsky views competence as an idealization abstracted away from the full range of linguistic behavior. As such, it deserves as much consideration as any idealization in science: if it yields interesting generalizations it is worthwhile. Still, one can make a distinction between 'soft' and 'hard' idealizations. A 'soft' idealization is acknowledged to be a matter of convenience, and one hopes eventually to find a natural way to re-integrate excluded factors. A standard example is the fiction of a frictionless plane in physics, which yields important generalizations about forces and energy. But one aspires eventually to go beyond the idealization and integrate friction into the picture. By contrast, a 'hard' idealization denies the need to go beyond itself; in the end it cuts itself off from the possibility of integration into a larger context.
> It is my unfortunate impression that, over the years, Chomsky's articulation of the competence/performance distinction has moved from relatively soft [. . .] to considerably harder. [**99**, p. 33]

Jackendoff suggests we adopt a 'soft' competence/performance distinction, adding a third component to the framework [**99**]. The theory of competence is seen as the characterization of phonologic, syntactic and semantic data structures as they are processed and stored in the brain of speakers during language acquisition. The theory of performance is the description of the actual occurrence of such data structures in language comprehension and production. The theory of neural instantiation is an account in terms of brain structures and processes of competence and performance. Jackendoff provides an architecture in which competence components (phonology, syntax and semantics, plus interface rules) interact in a manner that is consistent with the incrementality and the 'opportunism' (his label for immediacy) of language processing [**101**]. However, to solve the dilemma described above, it is not enough to show that competence determines the state-space available to users of a language during performance [**99**, p. 56]. The issue is, rather, whether there is interplay between competence and performance, that is – turning Jackendoff's tag line upside down – whether the logic of processing dictates the logic of competence, and to what extent.

As we saw above, in his early writings Chomsky claimed that theories of competence have nothing to learn from processing data [**23**]. Minimalists have suggested that syntax is adapted to the requirements holding at the interface with other cognitive modules, such as the sensory-motor and conceptual systems. However, they deny what functionalists on the contrary accept, namely that syntax is well-designed for use [**24, 87, 54**]. Evidence against adaptation to performance is provided, according to minimalists, by memory limitations, constructions such as garden-path and center embedding sentences, and so forth. Here two remarks are in order. The first is that such phenomena do not constitute evidence against adaptation *per se*, but rather (if anything like that is possible) against 'perfect' adaptation. Minimalists seem to commit what optimality theorists have called the 'fallacy of perfection' [**135**], consisting in confusing optimal outcomes, which are the result of equilibria between different variables, with best possible outcomes for just a subset of the factors involved, for instance the absence of unstable or ambiguous constructions (see [**164**] for discussion). The second remark is that, even if we assume that competence is neither perfectly nor optimally adapted to use, it still seems conceivable that performance constraints shaped competence rules. Therefore, the problem is not whether language is an adaptation – that some traits of competence reflect the outcomes of adaptive evolutionary processes acting on actual brain systems, including adaptation to communication needs, seems to be a widely accepted view [**87, 164, 54**]. The problem is rather: (how) can we construct a methodological framework in which it is possible to determine what aspects of competence can be explained adaptively?

The reason why generative linguistics does not seem capable of addressing this issue is, in our opinion, to be attributed more to how performance is defined than to a rigid view of the competence/performance distinction. Jackendoff [**99**, p. 30] rightly observes that in Chomsky's original proposal a large and heterogeneous set of phenomena were collapsed into 'performance': errors, shifts of attention, memory limitations, processing mechanisms, and so on. Only a very superficial assessment of the factors involved in language use could justify the notion that a single, relatively compact theory of performance could account for all those phenomena. It seems more reasonable to assume that different theories, developed using different analytical approaches, are necessary to understand how language interacts with memory and attention, how errors of different type and origin are produced (for also language disorders give rise to performance failures), and so on. We agree with Jackendoff on the characterization of competence and neural implementation, but we believe a more appropriate intermediate level should be chosen.

2.4. Marr's three-level scheme as applied to language. Jackendoff's updated view of the competence/performance distinction as a soft methodological separation, plus a theory of neural realization, resembles Marr's [**131**] tripartite scheme for the analysis of cognitive systems [**189**]. Marr suggested that cognitive processes should be modeled at three, nearly independent levels of analysis: a computational level (what is computed?), an algorithmic level (how is computation carried out?), and a level of physical implementation (what are the properties of the real machines that can execute the algorithms?). From this perspective, Jackendoff's trajectory away from Chomsky appears incomplete. So it may be worth asking what are the advantages of taking one step further: replacing the competence/performance distinction with Marr's scheme.

An important consequence of this choice is that performance theory is now seen as an intermediate level of analysis at which the algorithms and memory mechanisms supporting specific linguistic computations are described. That might seem a rather abrupt move, as it restricts the scope of performance to algorithms, and thereby leaves aside a large number of phenomena which, some might suggest, cannot be adequately treated in algorithmic terms. For instance, conscious inner speech is an important function of language [**18, 19**], and one in which there seems to be no definite input-output mapping involved. On the other hand, for those phenomena that are best treated as structured input-output processes, for example language production and comprehension, Marr's framework allows competence theories, if properly construed, to be investigated as part of actual information processing systems. Applications of this idea to semantics will be shown below and in chapters 2-3.

Does our appeal to Marr's three-level scheme solve the problems associated with the competence/performance distinction? It seems it does, because the variety of phenomena that were collapsed into performance can now be understood in their distinctive features.

For instance, working memory as involved in a given task can be examined at the level of algorithms. The algorithmic analysis may suggest a description of the memory architecture and the memory resources required by the system, and this constitutes a first step toward an explanation in neural terms. Conversely, memory organization constrains the classes of algorithms that can be computed by that machine, and the type of data structures that the computational theory can assume. An example of bottom-up adjustment is Yngve's **[223]** explanation of the dominance of right-branching over left-branching and center-embedding structures. Another example are 'minimal models' of discourse, as we shall see later.

Let us now apply this tentative methodological sketch to the semantics of tense, aspect and event structure. Our goal is to devise semantic analyses that are formally specified and cognitively motivated, that is, highlighting connections between the meanings of temporal expressions, planning and reasoning. The semantic analyses should also be algorithmically explicit, such that processing predictions, or general constraints on processing architecture, can be formulated. We hope to show that our theory of tense, aspect, and event structure not only meets these requirements, but can also be used to provide alternative explanations of existing experimental data on language comprehension. In chapters 2-6, this approach will be applied in more detail to other linguistic phenomena. Moreover, our own experimental data will be collected and used to test the processing hypotheses suggested by the semantic theory. The last part of this chapter puts our enterprise into a wider neuroscience-oriented perspective, introducing the 'binding problem for semantics'.

3. Planning, reasoning, meaning

3.1. The cognitive substrates of tense and aspect. We see it as the essential purpose of tense and aspect to facilitate the computation of event structure as described in a narrative. One consequence of this characterization is that, contrary to what generative and formal semantic approaches maintain, it is not very useful to study tense and aspect at the sentence level only. Tense, aspect and event structure really come into their own only at the discourse level **[26, 27]**. Tense and aspect, however, cannot by themselves determine event structure, and must recruit world knowledge. Examples (7a-c) will make clear what we have in mind.

French has several past tenses (Passé Simple, Imparfait, Passé Composé), which differ in their aspectual contributions. The following mini-discourses in French all consist of one sentence in the Imparfait and one in the Passé Simple. However, the structure of the set of events differs in each case.

(7) a. Il faisait chaud. Jean ôta sa veste. (Imp, PS)
It was hot. Jean took off his sweater.
b. Jean attrapa une contravention. Il roulait trop vite. (PS, Imp)
Jean got a ticket. He was driving too fast.
c. Jean appuya sur l'interrupteur. La lumière l'éblouissait. (PS, Imp)
Jean pushed the button. The light blinded him.

In the first example, the Imp-sentence describes the background against which the event described by the PS-sentence occurs. In the second example, the PS-event terminates the Imp-event, whereas in the third one the relation is rather one of initiation. These discourses indicate that world knowledge in the form of knowledge of causal relations is an essential ingredient in determining event structure. This knowledge is mostly applied automatically, but may also be consciously recruited if the automatic processing leaves the event structure underdetermined. It is the task of cognitive science to determine what this algorithm looks like, and how it is actually implemented. The prominent role of causal relationships is (7a-c) suggests a direction in which to look for that algorithm.

3.2. Planning, causality and the ordering of events. Stated bluntly, our hypothesis is:

> The ability to automatically derive the discourse model determined by a narrative is subserved by the ability to compute plans to achieve a goal.

Here we present several converging lines of evidence which lend some plausibility to this conjecture. Firstly, one distinguishing feature of human cognition is that it is goal-oriented, with goals that range from very short-term (get a glass of water) to very long-term (having sufficient income after retirement). In each case, the goal is accompanied by a plan which produces an ordered collection of actions, be they motor actions or transfers of money to a special account. The two cases differ in that the former plan is for the most part generated automatically, whereas the latter may involve a good deal of deliberation. More precisely, planning consists in

> the construction of a *sequence*[5] of actions which will achieve a given goal, taking into account properties of the world and the agent, and also events that might occur in the world.

In the literature, there have been numerous attempts to link the language capacity with the planning capacity. The setting is usually a discussion of the origins of language. Even if it is granted that some non-human primates have learned a primitive form of language, there is still a striking difference in language proficiency between chimpanzees and ourselves. It is still a matter of ongoing debate to determine exactly what this difference consists in. Some would say that the difference is in the syntax: human syntax is recursive, the chimpanzee's syntax (if that is the word) is not. One may then point to an analogy between language and planning. Language production can be characterized as transforming a semantic structure, to which the notion of linearity may not be applicable, into linear form, that is an utterance. Planning also involves linearization, and that is how the language-planning connection is drawn. An alternative strategy, not inconsistent with the first, is to show that the recursive structure of syntax is linked to the recursive structure (or hierarchical organization) of plans [**67, 191**]. Non-human primates engage in planning for time spans not exceeding 48 hours, as is known since Kohler's 1925 observations [**113**]. This has also been attested in squirrel monkeys in experiments by McGonigle et al. [**138**]. So complex planning sets humans apart from non-human primates, and it probably is one factor in explaining the human linguistic divergence with respect to other species.

A more direct path between language and planning was investigated experimentally by Trabasso and Stein [**202**] in a paper whose title sums up their program: *Using goal-plan knowledge to merge the past with the present and the future in narrating events on-line*. Trabasso and Stein argue that "the plan unites the past (a desired state) with the present (an attempt) and the future (the attainment of that state)" [**202**, p. 322], "[c]ausality and planning provide the medium through which the past is glued to the present and future" [**202**, p. 347]. They present the results of a study in which children and adults were asked to narrate a sequence of 24 scenes in a picture storybook called *Frog, where are you?*, in which a boy tries to find his pet frog which has escaped from its jar.[6] The drawings depict various failed attempts, until the boy finds his frog by accident. The aim of the study is to determine what linguistic devices, in particular temporal expressions, children use to narrate the story as a function of age. The authors provide some protocols which show a child of age 3 narrating the story in a tenseless fashion, describing a sequence of objects and actions without relating them to other objects and actions. None of the encoded actions is relevant to the boy's ultimate goal. Temporal sequencing comes at age 4, and now some of the encoded actions are relevant to the goal. Explicit awareness that a particular action is instrumental towards the goal shows up at age 5. At age 9, action-goal relationships are marked increasingly, and (normal) adults structure the narrative completely as a series of failed or successful attempts to reach the goal. We can see from this that there is some relation between children's developing sense of time and their ability to structure the narrative as the execution of a plan toward the goal of finding the frog: the child of age 3 seems glued to the present; the child of 4 is capable of

[5] More complex plans are possible which may involve overlapping actions.

[6] This is a classic experimental paradigm for investigating the acquisition of temporal notions in children. See Berman and Slobin [**6**] for methods, results and, last but not least, the frog pictures themselves.

including causal relations between events, states of mind and actions; these causal relations implicitly drive the narrative forward; the child of 5 can move from narrating some ongoing action to mentioning a goal state to be attained in the future and back again. The following quote suggests that there must be a *gradual* development of these capabilities:

> Inferring goals and plans involves considerable social and personal knowledge and places heavy demands on a narrator's working memory. The child who narrates events needs to attend to and maintain the current event in working memory; to activate and retrieve prior knowledge relevant to events, either in general or from earlier parts of the story, in order to interpret and explain the current event; and to integrate these interpretations into a context within a plan, all within the limitations of knowledge and working memory. In effect, over time the child is engaged in dynamic thinking, actively constructing and evaluating models and hypotheses about what is occurring. In so doing, the child creates a changing mental model that results in a long-term memory representation of what has occurred. [**202**, p. 327]

Working memory is thus essentially involved in this process of discourse integration, and failures in its operation may show up as deficiencies in the use of temporal expressions.

What is interesting for our purposes is that the ingredients that jointly enable planning have a prominent role to play in the construction of a discourse model. Take for instance causality, shown to be involved in the interpretation of (7a-c). Planning essentially requires knowledge of the causal effects of actions as well as of the causal effects of possible events in the world. Accordingly, the planning capacity must have devised ways of retrieving such knowledge from memory. Planning also essentially involves ordering actions with respect to each other and to events occurring in the world which are not dependent upon the agent. Furthermore, the resulting structure must be held in memory at least until the desired goal is attained. The reader can easily envisage this by considering the planning steps that lead to a pile of pancakes. For instance, causal knowledge dictates that one has to pour oil in the frying-pan before putting in the batter, and this knowledge has to remain active as long as one is not finished.

The fundamental *logical* connection between discourse processing and planning is that both are non-monotonic. When we plan, deliberately or automatically, we do so in virtue of our best guess about the world in which we have to execute our plan. We may plan for what to do if we miss the bus, but we don't plan for what to do if the bus doesn't come because the gravitational constant changes, even though that is a logical possibility. Similarly, the computation of a discourse structure may be non-monotonic. For instance, the reader who sees (8a) is likely to infer (that is, to read off from the discourse model) that Bill is no longer a member, but that implicature can easily be canceled, as in (8b):

(8) a. Bill used to be a member of a subversive organization.
 b. Bill used to be a member of a subversive organization, and he still is.

In cognitive terms, planning is part of 'executive function', an umbrella term for processes responsible for higher-level action control which are necessary for maintaining a goal and achieving it in possibly adverse circumstances. Executive function comprises maintaining a goal, planning, inhibition, coordination and control of action sequences. Hence our main hypothesis can be reformulated as follows: several components of executive function are involved in comprehension and production of tense and aspect. A corollary is that failures of executive function can show up in deviant use of tense and aspect and in impairments in processing temporal discourse, for instance in ASD (Autistic Spectrum Disorder), ADHD (Attention Deficit Hyperactivity Disorder, see section 3.2.2), and schizophrenia.

The link between planning and temporal semantics is provided by the notion of *goal*. In both comprehension and production, the goal is to introduce the event corresponding to the tensed VP into the event structure. This goal must have two components:

1. location of event in time;
2. meshing it with other events.

An example will make this clearer. Consider what goes on in comprehending

(9) Max fell. John pushed him.

On one prominent reading, the event described in the second sentence precedes, indeed causes, that described in the first sentence. The relevant goals are in this case:

update discourse with past event $e_1 = \mathtt{fall}(m)$ and fit e_1 in context
update discourse with past event $e_2 = \mathtt{push}(j,m)$ and fit e_2 in context

Planning must determine the order of e_1 and e_2, and to do so the planning system recruits causal knowledge as well as the principle that causes precede effects.

To give the reader a detailed picture of what goes on in such computations, we have to introduce some notation, borrowed from the Event Calculus **[212]**, which will also be useful for our discussion of experiments on language processing and the 'binding problem' later in this chapter. We make a distinction between *events* (denoted $e, e', ..., e_0, ...$) and *processes* or *fluents* (denoted $f, f', ..., f_0, ...$). We say that events occur or *happen* at a particular time, and represent this by the expression $Happens(e,t)$. By contrast, processes do not occur, but are going on at a particular time, and for this we use the predicate $HoldsAt(f,t)$. Events and processes can stand in causal relations. For instance, an event may kick off a process: $Initiates(e,f,t)$; or it may end one: $Terminates(e,f,t)$. We will use these predicates to mean the causal relation only. It is not implied that e actually occurs. Finally, a useful predicate is $Clipped(s,f,t)$, which says that between times s and t an event occurs which ends the process f. The predicates just introduced are related by axioms, of which we shall see a glimpse below. With this notation, and using $?\varphi(x)$ `succeeds` to abbreviate: 'make it the case in our discourse model that $\varphi(x)$',[7] we can write the two update instructions involved in comprehending the discourse as:

(10) $?Happens(e_1,t), t < now, Happens(e',t')$ `succeeds`

(11) $?Happens(e_2,s), s < now, Happens(e'',t'')$ `succeeds`

Here e' and e'' are variables for event types in the discourse context which have to be found out by substitution or, more precisely, *unification*. These two update instructions have to be executed so that $e'' = e_1$ and $s < t''$. If 'Max fell' is the first sentence of the discourse, we may disregard e'.[8] In order to formulate the causal knowledge relevant to the execution of these instructions, we introduce a process f (*falling*) corresponding to the event $e_1 = fall(m)$, where f, e_1 and e_2 are related by the following statements:

(12) $HoldsAt(f,t) \rightarrow Happens(e_1,t)$

(13) $Initiates(e_2,f,s)$

The system processing the discourse will first satisfy the update request corresponding to 'Max fell' by locating the event e_1 in the past of the moment of speech. The second sentence, 'John pushed him', is represented by the request (11) which contains the variable e''. The system will try to satisfy the goal by reducing it using relevant causal knowledge. Applying (12) and unifying[9] $e'' = e_1 = fall(m)$, the second request (11) is reduced to:

(14) $?Happens(e_2,s), s < now, Happens(e_1,t''), HoldsAt(f,t'')$ `succeeds`

The system now applies a general causal principle, known as *inertia*, which says that, if an event e has kicked off a process f at time t, and nothing happens to terminate the process between t and t', then f is still going on at t'. This principle rules out spontaneous changes, that is changes which are not caused by occurrences of events. Inertia can be formulated as the following axiom:

[7] This notation derives from logic programming. By itself, $?\varphi(x)$ denotes a *goal* or *query*, a request for a value a of x such that $\varphi(a)$ is true. The answer may be negative, if the database against which $\varphi(x)$ is checked contains no such individual. By $?\varphi(x)$ `succeeds` we mean that in such cases the database must be updated with an a making φ true. These instructions or requests for updates are also known as *integrity constraints*.

[8] Here we regard context as provided by the preceding discourse, but one may conceive of 'forward-looking' notions of context as well.

[9] This form of unification will be important in our discussion of the 'binding problem' for language.

(15) $Happens(e,t) \wedge Initiates(e,f,t) \wedge\ t < t' \wedge \neg Clipped(t,f,t') \rightarrow HoldsAt(f,t')$

Using this axiom, the request (14) is further reduced to:

(16) $?Happens(e_2,s), s < now, Happens(e_1,t''), Happens(e,t), Initiates(e,f,t), t < t''$, $\neg Clipped(t,f,t'')$ `succeeds`

Using (13) and unifying $e = e_2 = push(j,m)$ and $s = t$ we reduce this request to:

(17) $?Happens(e_2,s), s < now, Happens(e_1,t''), s < t'', \neg Clipped(s,f,t'')$ `succeeds`

This is a definite update request which almost says that *push* precedes *fall*, except for the formula $\neg Clipped(s,f,t'')$, which expresses that f has not been terminated between s and t''. If f were terminated between s and t'', we would have a situation as in:

(18) Max fell. John pushed him a second time and Max fell all the way to the bottom of the pit.

Since we have no positive information to this effect, we may assume $\neg Clipped(s,f,t'')$. This form of argument is also known as *closed world reasoning*: 'assume all those propositions to be false which you have no reason to assume to be true'. Closed world reasoning is essential to planning, and to discourse comprehension, as it allows one to discount events which are logically possible but in practice irrelevant. The final update request is thus:

(19) $?Happens(e_2,s), s < now, Happens(e_1,t''), s < t''$ `succeeds`

which is the instruction to update the discourse model with the past events e_1 and e_2 such that e_2 precedes e_1.

Just as plans may have to be revised in mid-execution (for instance, if it turns out there is not sufficient oil to produce the projected number of pancakes), discourse models may have to be recomputed when additional information is provided. Suppose the discourse does not stop after 'John pushed him' but, instead, continues:

(20) Max fell. John pushed him, or rather what was left of him, over the edge.

One obvious interpretation is that now $e_2 = push(j,m)$ comes after $e_1 = fall(m)$. This is the result of a recomputation, since after the first 'him' the hearer may have inferred that e_2 precedes e_1. Let us give a brief, informal sketch of this recomputation. The phrase 'or rather what was left of him' suggests Max is now *dead*, therefore the update request corresponding to the second sentence is something like:

(21) $?Happens(e_2,s), s < now, HoldsAt(dead(m),s), Happens(e'',t'')$ `succeeds`

perhaps together with a requirement to the effect that the entire pushing event occurs while *dead*(*m*) obtains. It now seems reasonable to assume that, at the start of *falling* (the process denoted by f), Max is still *alive*. Unifying $e'' = e_1$ and applying property (12), the request reduces to finding instants s, t'' such that:

(22) $?Happens(e_2,s), s < now, HoldsAt(dead(m),s), HoldsAt(alive(m),t'')$, $Happens(e_1,t'')$ `succeeds`

can be satisfied. Since *alive* always precedes *dead* and not conversely, it follows that we must have that $e_1 = fall$ precedes $e_2 = push$.

In summary, what we have outlined here is a universal computational mechanism for determining event structure based on planning. Temporal expressions (not only tense and aspect, but also temporal connectives as will be seen in section 3.2.3) are hypothesized to determine requests to be satisfied by an update of the current discourse model. Processing these requests involves unification, search through semantic memory, as well as setting up temporary structures in working memory. These computations might leave characteristic traces on neurophysiological responses like event-related potentials (ERPs), a hypothesis to which we will return later.

3.2.1. *Computing event structures for (PS, Imp) combinations.* Similar arguments apply to the French examples with which we started this section:

(7) a. Il faisait chaud. Jean ôta sa veste. (Imp, PS)
It was hot. Jean took off his sweater.

Intuitively, this narrative determines an event structure in which *hot* acts as a background which is true all the time, and the foregrounded event (taking off one's sweater) is located within this background. One arrives at this structure by means of the following argument. World knowledge contains no causal link to the effect that taking off one's sweater changes the temperature. The goal corresponding to the first sentence dictates that it is hot at some time t before *now*. By the principle of inertia, the state *hot* must either hold initially (at the beginning of the narrative) or have been initiated. The latter requires the occurrence of an initiating event, which is however not given by the discourse. Therefore, *hot* holds initially. Similarly, no terminating event is mentioned, so *hot* extends indefinitely, and it follows that the event described by the second sentence must be positioned inside *hot*.

The second example dates back to the bygone days when speeding cars were stopped by the police instead of being photographed:

(7) b. Jean attrapa une contravention. Il roulait trop vite. (PS, Imp)
Jean got a ticket. He was driving too fast.

It is given that the event of getting a ticket occurred sometime in the past, and it is also given that the fluent *speeding* was true sometime in the past. Hence, it holds initially or has been initiated. We have to determine the relative position of event and fluent. World knowledge yields that getting a *ticket* terminates, but does not initiate *speeding*. Because this is the only event mentioned, *speeding* holds from the beginning of discourse, and is not reinitiated once it has been terminated.

In the third example, the same order of the tenses yields a different event order, guided by the application of causal knowledge:

(7) c. Jean appuya sur l'interrupteur. La lumière l'éblouissait. (PS, Imp).
Jean pushed the button. The light blinded him.

One (occurrence of an) action is mentioned, pushing the light button, which has the causal effect of initiating the light being on when its current state is off. No terminating event is mentioned, therefore the light remains on. It also follows that the light must be off for some time prior to being switched on, and that it must be off at the beginning of discourse. The definite article in '*La* lumière' leads to a search for an antecedently introduced light, which successfully terminates after unification with the light introduced in the first sentence. As a consequence, it is *this* light which is too bright.

3.2.2. *Deviant verb tenses and ADHD.* The preceding considerations can be applied to those psychiatric disorders that are known to involve language impairments, sometimes of a rather subtle kind. For instance, children with ADHD have trouble with retelling a story, a task that involves presenting information so that it is organized, (temporally) coherent, and adjusted to the needs of the listener. The ability to attend to these requirements presupposes that one is able to retain goals in working memory while planning the necessary steps and monitoring their execution. This ability requires executive function as defined above [**170**].

The theory presented here can be used to constrain predictions on deviant narration in children with ADHD [**214**], who appear to have difficulty in maintaining goals in working memory [**65**]. Recall that update requests, that is the goals to be satisfied, corresponding to a VP's tense and aspect, consist of two components:

1. location of an event in time;
2. meshing the event with other events.

If a child has trouble maintaining a goal in working memory, this may lead to a simplified representation of that goal. In the case of verb tenses, the most probable simplification is of 'location of event in time' (never mind the meshing with other events), since this involves the least processing (search through semantic memory and unification). This simplification affects both comprehension and production, the case of interest here. Indeed, in producing speech which is attuned to the needs of the listener, the speaker may construct a discourse model of his own utterances, to determine whether it is sufficiently unambiguous. Now, a typical finding in this area is summarized in the following quote by Purvis et al.:

> ADHD groups exhibited a higher frequency of sequence errors, which reflects a breakdown in the global organization of story theme. The pure-ADHD group [i.e. the group without reading disability] had difficulties in local organization, reflected in ambiguous references [of pronouns referring to persons or events]. These are failures in cohesion which make it difficult for the listener to follow the speaker's train of thought. Ambiguous references can result from a failure to organize and monitor the cohesion of sentences, as well as from a failure to take into account the needs of the listener. [**170**, p. 141]

This seems to suggest that there is indeed a connection between language comprehension and production, and the processing of goals.

3.2.3. *Temporal connectives.* Processing considerations suggested by the Event Calculus can be used as alternative explanations of existing ERP data. Here is an example. Münte et al. [**147**] recorded ERPs while participants read sentences that differed only in the temporal connective, for instance:

(23) a. After the scientist submitted the paper, the journal changed its policy.
b. Before the scientist submitted the paper, the journal changed its policy.

An additional test assessing participants' working memory span was carried out. Subjects read aloud increasingly longer sequences of sentences. After each sequence, they attempted to recall the last word of each sentence in the order in which these were presented. Scores were then used to group subjects – high, medium and low working memory span. 'Before' sentences elicited a larger sustained negativity. The amplitude was maximal at left anterior electrodes, where ERP responses to 'before' and 'after' diverged around 300 ms after word onset. The effect lasted throughout the sentence, and was largest during the second clause. The ERP difference between 'before' and 'after' sentences was positively correlated with working memory spans: the higher the working memory scores, the larger the ERP effect.

According to Münte et al., the observed pattern of effects can be explained as follows. As reported above, ERP waveforms diverged within 300 ms following sentence onset, that is soon after the temporal connective was encountered. 'After' and 'before' rapidly activate different chunks of conceptual knowledge stored in memory: in sentence-initial position, 'after' indicates that the events are presented in their actual order; 'before' indicates that the order is reversed. As regards the larger sustained negativity elicited by 'before' sentences, it is hypothesized that these require additional computations as a consequence of presenting the events in reverse temporal order.

An important assumption, upon which the account of Münte et al. ultimately rests, is that the order in which the events are mentioned in discourse is the only difference between 'after' and 'before' sentences. This follows from their claim [**147**, p. 71] that what is encoded into the meanings of 'after' and 'before' are relations of temporal order between events, that is temporal precedence and temporal consequence. A review of the linguistic literature on the subject, however, suggests that there is more to the semantics of temporal connectives than event order. There is indeed an asymmetry between the inferences licensed by 'after' and 'before'. While 'after' always entails that the subordinate clause is true, 'before' allows for the subordinate clause to be either true or false, depending on the information provided by the main clause [**3, 29**]. (23b) suggest that the content of a 'before' clause, if the earlier event is likely to prevent the later one from occurring, will be interpreted as false or, more precisely, as non-veridical. In the situation described by (23b), the scientist might have not been able to submit the paper precisely because of the journal's policy change. The reading in which the subordinate is interpreted as false is still consistent with the overall truth of the sentence, hence non-veridical readings seem available to subjects.

Our analysis is based on the mathematical theory of tense and aspect proposed by van Lambalgen and Hamm [**212**], which has been briefly described and applied in sections 3.1 and 3.2. In that framework, tense, aspect and temporal adjuncts are represented by integrity constraints. Recall that these are instructions in the form of obligations and prohibitions to update the current model such that a representation making discourse true is incrementally computed. 'After' sentences must satisfy three integrity constraints. Two are introduced by the subordinate clause and require an update of the unification space such that the relevant activity (the submission process) holds in the past of the moment of speech and is finished before t'', where the event occurring at t'' is to be specified by the subsequent main clause:

(24) a. $?HoldsAt(submit,t) \wedge t < now$ `succeeds`
b. $?Happens(finish,t'), t' < t'' < now, Happens(e',t'')$ `succeeds`

The variable e' denotes an event type, and has to be unified with the event described by the main clause. Here *finish* is the culminating event and is followed by the goal state, that is a complete submission. The main clause introduces the second integrity constraint:

(25) $?Happens(change,t''), t'' < now$ `succeeds`

where *change* denotes the journal's policy change. The unification $e' = change$ leads to

(26) $?Happens(finish,t')$, $t' < t'' < now$, $Happens(change,t'')$ `succeeds`

This is the only binding operation required by 'after' sentences. Condition (26) implies that 'after' licenses veridical readings only.

'Before' sentences introduce three analogous conditions. The first two are added to the unification space by the subordinate clause. The second is derived from (24b) by replacing the obligation with a prohibition:

(27) a. $?HoldsAt(submit,t) \wedge t < now$ `succeeds`
b. $?Happens(finish,t'), t' \leq t'' < now, Happens(e',t'')$ `fails`

Also in this case e' is an event type to be unified with a token furnished by the main clause. The third condition introduced by the main clause is identical to its 'after' counterpart (25). The system will also in this case unify $e' = change$, reducing (27b) to the request

(28) $?Happens(finish,t'), t' \leq t'' < now, Happens(change,t'')$ `fails`

Notice that (28) can be satisfied in different models, because the integrity constraint fails if and only if *at least one* of its conjuncts fails. Suppose there is no information in declarative memory to the effect that a journal's policy change is a likely obstacle toward a submission. Closed world reasoning yields $\neg Terminates(change,submit,t)$. From this, it can be concluded that the submission was completed, because the process took place as is required by (27a), and it can moreover be assumed to be finite. As both events (the complete submission and the policy change) took place, the only solution left for the system is to add $t'' < t' < now$ to the unification space in order to ensure that (28) fails. This corresponds to the veridical reading in which the submission follows the journal's policy change. Suppose however that $Terminates(change,submit,t)$ is retrieved from declarative memory. The submission process is then interrupted before it is completed. This implies that the culminating event did not take place, and that the goal state was not attained. The first conjunct of (28) fails, yielding a non-veridical reading.

In order to consider the implications of our analysis at the algorithmic level we would need the full computational machinery of the Event Calculus, combined with hypotheses on the time-course of linguistic processes, for instance immediacy [**133**]. Here we can only provide a qualitative description of the steps involved in interpreting temporal connectives. Processing 'after' and 'before' sentences amounts to incrementally constructing discourse models which satisfy (24) and (27). Processing 'after' clauses consists in computing a model in which (24a) and the first two conjuncts of (24b) are satisfied – our version of immediacy implies that processing algorithms can satisfy incrementally only the conjuncts of integrity constraints of the form $?\varphi(x)_1,...,\varphi(x)_n$ `succeeds`. Processing the main clause will involve forcing an update satisfying (25) and unifying event variables, which automatically leads to a model satisfying (26). Processing 'before' clauses starts with the computation of a model in which (27a) can be satisfied. At that stage, however, none of the conjuncts of (27b) can be satisfied, for the integrity constraint is of the form $?\varphi(x)_1,...,\varphi(x)_n$ `fails` and as such can lead to very different updates of the discourse model. The algorithm which should update the model according to the information provided by the subordinate clause must either wait for the main clause to be interpreted, or it must commit to a defeasible interpretation of the 'before' clause – a veridical one, assuming closed world reasoning applies. The main clause, and related causal knowledge retrieved from declarative memory, determines whether the first conjunct of (27b) can be satisfied and so the second must fail, as in a veridical reading, or whether it is the first conjunct that fails, as in a non-veridical reading. If the execution of these updates is delayed, (27b) must be maintained in working memory until relatively late before it can be evaluated. If a veridical reading is computed while processing the 'before' clause, a recomputation might take place when the subordinate clause is encountered. On both accounts, 'before' sentences are more taxing on working memory resources available for unification, especially during later processing stages when the interpretation of the two clauses has to be finalized in a coordinated manner.

This rather sketchy formal analysis shows that semantic notions can be mapped onto processing reality, and can provide alternative explanations of existing experimental data. This suggests there is a role for semantics in generating hypotheses for the psycholinguist and the cognitive neuroscientist to test. Obstacles to this kind of integration lie not so much in our limited knowledge of language and the brain, as in research methodologies positing strict separations between domains of inquiry, where in fact only differences between levels of analysis exist.

4. The binding problem for semantics

The goal of a theory of language is to deliver analyses at each of Marr's levels, and to bridge them in a perspicuous manner. One way of achieving this is to define a notion that acts as a 'wormhole' [**94**] connecting linguistic structures, algorithms, and neurobiological events. A candidate notion is that of 'unification', which has been applied on several occasions in this chapter and will be seen at work in later chapters too.

An influential statement of the 'binding problem' for cognitive representations is due to von der Malsburg [**219**], who regarded the binding approach to brain function as a response to the difficulties encountered by classical connectionist networks. Von der Malsburg [**220**] refers to a well-known example by Rosenblatt [**179**] to illustrate the issue. Let us consider a network for visual recognition constituted by four output neurons. Two neurons fire when a specific shape (either a triangle or a square) is presented and the other two fire depending on the shape's position (top or bottom of a rectangular frame). So, if there is a square at the top, the output will be [square, top]. If there is a triangle at the bottom, the output will read [triangle, bottom]. However, if a triangle and a square are presented simultaneously, say, the triangle at the top and the square at the bottom, the output would be [triangle, square, top, bottom], which is also obtained when the triangle is at the bottom and the square at the top. This is an instance of the 'binding problem'. Malsburg writes:

> The neural data structure does not provide for a means of binding the proposition top to the proposition triangle, or bottom to square, if that is the correct description. In a typographical system, this could easily be done by rearranging symbols and adding brackets: [(triangle, top),(square, bottom)]. The problem with the code of classical neural networks is that it provides neither for the equivalent of brackets nor for the rearrangement of symbols. This is a fundamental problem with the classical neural network code: it has no flexible means of constructing higher-level symbols by combining more elementary symbols. The difficulty is that simply coactivating the elementary symbols leads to binding ambiguity when more than one composite symbol is to be expressed. [**219**, p. 96][10]

Examples of the binding problem are bistable figures such as Necker's cube and Jastrow's duck-rabbit, where the exact same visual features of the stimulus lead to two incompatible representations, depending on how these features are bound together. Since the availability of different representations essentially depends upon the geometric properties of the figure,

[10] Different solutions to Rosenblatt's problem are possible. See von der Malsburg [**220**] for a proposal in line with the binding hypothesis and Riesenhuber and Poggio [**176**] for an alternative approach.

rather than upon the constitution of perceptual systems as would be the case, for example, for after images [**131**, pp. 25-26], bistability requires an explanation at Marr's computational level, where properties of stimuli are described and related to information processing goals. Without a characterization of the geometric properties of the figure, and of the mappings between the figure and the two different entities which it can stand for, there would be no basis upon which to claim that the two representations are mutually exclusive.

There exist analogous cases of structural ambiguity in language:

(29) a. The woman saw the man with the binoculars.
b. Respect remains.

Example (29a) has two alternative syntactic representations, one in which the phrase 'with the binoculars' is a PP attached to the NP 'the man' (the man that was seen by the woman had binoculars), and another in which it modifies the VP (the woman used binoculars to see the man). Here too the features of the stimulus lead to two interpretations, depending on which attachment option is eventually pursued. These sentences typically result in specific ERP effects, suggesting that syntactic binding is a genuine information processing problem for the brain. Sentence (29b) also has two possible parses, and this has consequences for its meaning: it can either be used as a directive speech act, if 'respect' is the verb and 'remains' the object noun; or it can be used as an assertion, if 'respect' is the object noun and 'remains' the verb. 'Before' sentences and accomplishments in the progressive (chapters 2-5) are also examples of semantic bistability, as they allow two mutually exclusive interpretations.

There are some similarities between perceptual bistability in the visual and linguistic domains, such as the fact that in both cases we seem to 'flip' between the two incompatible representations. But there is also a deeper analogy between the two: structural ambiguity is defined at the topmost level of analysis in both cases, as Marr pointed out [**131**, pp. 25-26]. Without an independent characterization it remains unclear why such representations are mutually exclusive in the first place. Extending Marr's line of argument, we emphasize that the binding problem for semantics is best formulated at the computational level, although attempted solutions are bound to require significant contributions at all levels of analysis, including – perhaps most interestingly – the level of neural implementation [**72, 73**].

There is a computational formulation and solution of the binding problem for syntax [**71**], which will be applied in chapter 4. However, a parallel piece of work for semantics is still missing. The purpose here is not to lay out a full-fledged theory of semantic unification. Rather, the three main principles of semantic unification will be examined in some depth, combining formal semantic analyses with ERP data as exemplified in section 3.2.3. Here is a summary of what the reader may expect from the remaining chapters:

Principle	Research question	Theory	Experiment
Immediacy	*How do syntactic and semantic unification interact in early stages of tense processing?*	ch. 4	ch. 4
Monotonicity	*Is semantic recomputation involved in processing progressive constructions?*	ch. 2	ch. 5
Compositionality	*Is enriched composition involved in processing complement coercions?*	ch. 3	ch. 6

CHAPTER 2

The processing consequences of monotonicity

This chapter is a modified version of G. Baggio & M. van Lambalgen. The processing consequences of the imperfective paradox. *Journal of Semantics* 24, 2007: 307-330.

1. Introduction

Recently, a number of studies have brought experimental data to bear on semantic theories (see, for instance, the work on quantifiers by Geurts and van der Slik [**64**] and McMillan et al. [**140, 141**]). However, formal semantics and psycholinguistics have reached their most important results independently, even when the issues at stake could have been addressed more completely by joining efforts. One example is the study of discourse-based inference in psycholinguistics, where formal notions of truth, entailment and veridicality have often been neglected (see Cook et al. [**28**] and Frank et al. [**57**] among others).

One assumption behind this book is that there exists a relatively unexplored territory in which the two disciplines can interact productively. Here we consider a small portion of this territory: monotonicity and its processing consequences. We first apply Marr's scheme to semantics, and we introduce the principle of monotonicity in discourse processing. Next we present the linguistic phenomenon which we shall be concerned with: the imperfective paradox. In section 2, we provide an analysis of the paradox based on the Event Calculus (EC) [**212**], a planning formalism characterizing a class of models that can be computed by connectionist networks. In section 3, we report the results of a questionnaire that supports the semantic theory and suggests that different aspectual classes of past progressive verbs yield different entailment patterns. In section 4, a processing model is outlined combining the Event Calculus with the psycholinguistic principle of immediacy in the framework of recurrent networks. The model is used to derive predictions concerning the ERP correlates of the computations described by the Event Calculus.

1.1. Methodological preamble. From a semantic analysis of a linguistic structure it is usually not possible to derive predictions concerning the complexity and time-course of the processes involved in producing and comprehending utterances in which the structure occurs. There may be cases in which processing hypotheses can be formulated on the basis of the semantic theory alone [**64, 140**], but this approach is unlikely to work in general. The reason is that processing depends upon the particular algorithms and neural mechanisms that, in a physical system such as the human brain, compute the linguistic structures posited by the theory – or functionally equivalent ones. Since semantics typically does not describe algorithms and neural mechanisms (nor perhaps it should), there is no direct way to relate semantic theories to what is observed in language processing experiments.

The solution is to adopt a theoretical framework that allows formal semantic analyses to be explicitly related to processing algorithms inspired by the available psycholinguistic evidence, and ultimately to computational models of the kind investigated in neuroscience. Marr's [**131**] scheme for the analysis of cognitive systems seems an appropriate choice in this regard. Some motivation for this choice was provided in chapter 1.

Marr argued that information-processing systems should be understood at three nearly independent levels of analysis. The first level contains a description of the computations to be performed by the system and, more precisely, a characterization of the *goals* that have to be attained in order to solve a particular information-processing problem. For example, a sentence *S* can be seen as introducing a specific goal, namely the construction of a cognitive model making *S* true. In the semantic analysis proposed below, the tense and aspect of VPs like the past progressive are treated as instructions to update the current discourse model so as to achieve that goal. We see it as the task of semantics to describe information-processing goals and update instructions. The actual processing steps are described at the intermediate level, where constraint-satisfaction algorithms, implemented in artificial neural networks [**132**], and processing principles such as immediacy and incrementality are combined. Marr adds a third level, at which the neurobiological architecture is described. We discuss issues relevant for the computational level in section 2 and for the algorithmic level in section 4.

1.2. Estimating meaning: beyond monotonicity. Several basic facts about computing discourse models are established in psycholinguistic research: semantic representations are built up incrementally in close temporal contiguity with the input signal [**133**]; there is no processing distinction (in terms of time-courses and type of neurophysiological responses) between sentence- and discourse-level integration of word meanings [**208**]. However, some fundamental issues regarding discourse processing remain unanswered. For instance: how much information does a model contain, and is the revision of its content allowed?

We consider discourse models as 'minimal' in the sense that, information which is not provided by what is said or cannot be inferred from it, is (temporarily) assumed to be false. For example (inspired by Lascarides and Asher [**122**]; recall also the formal analysis given in chapter 1), upon hearing

(1) Max fell.

a minimal model is computed in which Max fell and nothing else is the case. If the narrative continues as in

(2) Max fell. John pushed him.

a new model is computed based on the lexical meanings of 'John' and 'pushed', and world knowledge to the effect that, barring unforeseen circumstances, pushing causes falling. The initial minimal model is *extended* with a representation of the pushing event as preceding the falling event.

Incrementality means that processing 'shadows' the input signal, but *not* that models are built up monotonically, without the need for revision. Precisely because interpretation is incremental, and one has to face incomplete or uncertain information at some processing stages, minimal models are *estimates* of the meaning intended by the speaker. For instance, upon hearing (2), we may conclude that John's pushing preceded Max's falling, the former event being the cause of the latter. However, if (2) were to continue, as in

(3) Max fell. John pushed him, or rather what was left of him, over the edge.

we would infer, on the contrary, that John's pushing is part of an effort to dispose of Max's corpse, reversing the order of events suggested by (2).

Inferences that can be withdrawn based on further information are called *non-monotonic*. Classical monotonic logic is assumed by most proposals in formal semantics: interpretation, also when it is regarded as incremental, as in dynamic semantics [**107, 69**], proceeds via the construction of a monotonic chain of models, where each preserves the semantic structure of the preceding one. Still, non-monotonicity appears necessary in light of both linguistic considerations [**122, 124, 212**] and experimental data [**196**]. The imperfective paradox seems to be one such case in favor of non-monotonicity.

1.3. The imperfective paradox. Vendler [**218**] famously classified VPs as states ('know', 'love' etc.), activities ('write', 'run' etc.), accomplishments ('write a letter', 'run a mile' etc.) and achievements ('finish', 'reach' etc.). Here we shall deal with the inferences licensed by sentences containing activities and accomplishments in the past progressive. The following example involves the accomplishment 'write a letter':

(4) The girl was writing a letter when her friend spilled coffee on the tablecloth.

From (4) the reader would typically conclude that, barring unforeseen circumstances, the girl attained the desired goal and would thus assent to the statement 'The girl has written a letter'. Such an inference is based on the assumption that spilling coffee on the tablecloth is usually *neutral* with respect to the writing activity, that is, it is not a typical proximate cause leading to its termination. It is possible to imagine scenarios in which writing is temporarily interrupted or even terminated by the accident. Nonetheless, as the data reported in section 3 will demonstrate, failing to mention a disabling condition in the discourse is sufficient for leading the reader to assume that there was no such obstacle to attaining the intended goal.

Inferences to a goal state are *non-monotonic*, i.e. they can be suppressed if the discourse describes an event which terminates the relevant activity:

(5) The girl was writing a letter when her friend spilled coffee on the paper.

Assuming that writing was intended to occur on the same paper sheets on which coffee was spilled, the accident is sufficient for terminating the activity, and it is therefore a *disabling* condition for obtaining a complete letter. Accordingly, on the basis of (5) the reader would be more likely to assent to 'The girl has written no letter'.

One important observation is that suppression can obtain only with accomplishments, and not with activities [**180**]. Since a sentence containing an activity in the past progressive, such as 'writing letters', does not involve a canonical goal, it is interpreted as entailing that 'The girl has written one or more letters' regardless of the consequences of the second event on the writing activity:

(6) The girl was writing letters when her friend spilled coffee on the tablecloth.
(7) The girl was writing letters when her friend spilled coffee on the paper.

Upon reflection, there is something paradoxical about examples (4) and (5) which is not found in (6) and (7). Even though it belongs to the meaning of the accomplishment 'writing a letter' that the writing activity is directed towards the goal state of the complete letter, the actual occurrence of that goal state can be denied without contradiction. However, how can a seemingly essential component of the meaning be denied without destroying the meaning itself? This is known as the 'imperfective paradox'. The semantic literature is replete with attempted solutions of the paradox, which range from explaining the problem away [**143**] to various invocations of possible worlds [**42, 120, 36**]. It is impossible to review all proposed solutions here; we will focus on representative claims instead. Possible worlds analyses are based on the idea that

> the progressive picks out a stage of a process/event which, if it does not continue in the real world, has a reasonable chance of continuing in some other possible world. [**36**]

They differ however in the (largely informal) descriptions of the possible worlds used. For example, Dowty [**42**] would claim that the following are equivalent:

a. 'The girl is writing a letter' is true in the actual world;
b. 'The girl will have written a letter' is true in all so-called 'inertia worlds', worlds which are identical with the present world until 'now', but then continue in a way most compatible with the history of the world until 'now'.

These approaches are intensional in the formal sense of using possible worlds. In fact, most authors (but not all) would agree that the progressive creates an intensional context: even though the accident in (5) may have terminated the writing activity at a stage in which it was unclear whether the girl was writing a letter or, say, a memo, still only one of

(8) The girl was writing a letter.
(9) The girl was writing a memo.

is true of the situation described by (5). Explicitly denying that the progressive creates an intensional context, Michaelis [**143**] argues that[1]

> the Progressive sentence 'She is drawing a circle' denotes a state which is a subpart not of the accomplishment type *She–draw a circle* but of the activity type which is entailed by the causal representation of the accomplishment type. Since this activity can be identified with the preparatory activity that circle drawing entails, circle drawing can in principle be distinguished from square drawing etc. within the narrow window afforded by the Progressive construal [and] does not require access to culmination points either in this world or a possible world.

We find this questionable, for without access to a person's intention it may be very hard to tell initially whether she is drawing a circle or a square, writing a letter or a memo. But that person's intention in performing an activity is characterized precisely by the associated consequent state, even though the latter cannot yet be inferred from the available data.

Here the Event Calculus will come to our rescue, because the notion of goal or intention is built into the semantics from the start. In particular, the meaning of a VP is represented by a *scenario* which describes a plan for reaching the goal. However, unlike approaches such as Parsons' [**157**] where one quantifies existentially over events, the scenario is a universal theory and does not posit the occurrence of the intended consequences. Although the plan is appropriate for that purpose, attaining the goal is guaranteed only in a *minimal model* (in which no unforeseen obstacles occur) of the scenario plus the axioms of the Event Calculus. The meaning of the accomplishment as is embodied in the scenario involves a culminating event *type* – which therefore must exist – but no existential claims are made concerning the corresponding event *token*, which, as in example (5), may also fail to occur. Type and token are handled by different mechanisms in the Event Calculus. These notions form the basis of our semantic analysis of the imperfective paradox, to which we now turn.

2. Goals and completion inferences

The reference to goal states in the preceding section suggests that a semantic analysis of the progressive can be based on a planning formalism which is able to talk about events and goals, and which includes a theory of causality together with a principle of inertia [**84**]. Such a formalism was developed by van Lambalgen and Hamm [**212**], and it consists of an Event Calculus which has found applications in robotics, reformulated using the computational machinery of Constraint Logic Programming.

The reader may wonder how can planning provide a source of inspiration to semantics. The reason can be found in the nature of computation in planning, which typically proceed as follows. First a goal is specified, which may be an action to be performed at a particular location – for instance, pick up outgoing mail in an office building. Next a plan is devised, that is a sequence of actions to get to the required location, derived by backward chaining from the goal to obtain a sequence of subgoals, the last one of which can be executed in the agent's initial position and state. Planning requires a situation model (including a map of the building, a causal theory of the agent's actions, a specification of values of variables such as 'door open/closed', the agent's initial position and state, a record of its past and current actions), a repertoire of possible actions ('follow wall', 'go through door') and observations ('door open/closed'). While the agent executes the plan, it also registers its observations and actions in the situation model. Information about its actions may be important for the agent to estimate its current position. Particularly relevant for our purposes is that a plan might have to be recomputed in mid-course when the initial situation model is updated due to new observations, for instance a closed door which was expected to be open on the basis of the initial model, or a wrong estimate of the agent's current position. Later on we will see how this 'recomputation' relates to the imperfective paradox.

[1]Replace 'drawing a circle' with 'writing a letter' and 'drawing a square' with 'writing a memo' in Michaelis' examples to see the connection with our examples.

This concise description should be sufficient to enable the reader to see the connection with language comprehension. The reader/hearer starts off with an initial discourse model in which an incoming sentence or clause must be integrated. Suppose the main verb of the sentence is non-stative, for example an activity. If the sentence is in one of the simple tenses, it is unpacked in the relevant action and its participants and the discourse model is updated accordingly. This is the analogue of updating the situation model with representations of individuals and actions in planning. In more complex cases, such as sentences involving accomplishments like (4) and (5), the VP is taken to express the existence of a goal-directed plan. If, on the contrary, the main verb of the sentence is stative, the sentence can be viewed as analogous to an observation report and the discourse model is updated with a property.

2.1. The Event Calculus. The Event Calculus is a planning formalism which allows one to talk about actions, goals and causal relations in the world. Its function is to return a plan given a goal, the initial state and causal relations. Formally, the Event Calculus is a many-sorted predicate logic. It has two different sorts for events, viewed either perfectively or imperfectively. The former are called *event types* and are symbolized by $e, e', ..., e_0, ...$. The latter are called *fluents* (Newton's name for time-dependent variables) and are symbolized by $f, f', ..., f_0, ...$ One may think of event types as action types, such as, for example, 'break' or 'ignite'. Fluents can be thought of as time-dependent properties, such as 'being broken' or 'writing'. The time parameter in fluents is implicit, but they can have further parameters, for instance for the subject of 'writing'. The real distinction between event types and fluents comes from the different roles they play in the axioms of the Event Calculus. The universe contains also sorts for individuals ('the girl'), real numbers interpreted as instants of time, and various other real quantities, such as position, velocity, or degree of some quality.

The primitive predicates comprise the minimum necessary to talk about two forms of causality, instantaneous (as in a collision) and continuous (as when a force is acting):

(10) a. *Initially*(f) : 'fluent f holds at the beginning of the discourse'
b. *Happens*(e,t) : 'event type e has a token at time t'
c. *Initiates*(e,f,t) : 'the effect of event type e at time t is the initiation of f'
d. *Terminates*(e,f,t) : 'the effect of event type e at time t is the termination of f'
e. *Trajectory*(f_1,t,f_2,d): 'if f_1 holds from t until $t+d$, then at $t+d$ f_2 holds'
f. *Releases*(e,f,t) : 'event type e releases the continuously-varying fluent f'[2]
g. *HoldsAt*(f,t) : 'fluent f holds at at time t'

While (10a-d) are the predicates required by instantaneous change, (10e-f) are used to model continuous change.

Three further notions will be deployed in the semantic analysis presented below. The first is that of *scenario*.[3] Whereas the axiom system EC provides a macro-theory of causality, scenarios provide micro-theories stating the specific causal relations obtaining in a given situation. Scenarios can be used to describe the causal environment of actions and events at the level of granularity expressed in natural language, such as for instance writing a letter or drawing a circle. Scenarios can be taken to represent in a tenseless fashion the meaning of VPs. For example, the scenario for the accomplishment 'writing a letter' specifies that the writing activity is causally related to the amount of letter produced, that the termination of the activity is temporally contiguous to the completion of the letter and so on.

The second notion is that of *integrity constraint*. As we said above, we regard a sentence as introducing an information-processing goal ('Make S true') to be achieved by updating

[2]The predicate *Releases* is used to reconcile the two notions of causality. While instantaneous change leads to one form of inertia, where properties do not change their value between the occurrences of two events, continuous change requires that variable quantities may change value without concomitant occurrences of events. The solution of this conflict is to exempt, by means of the predicate *Releases*, those properties which we want to vary continuously from the inertia of the first form of causation.

[3]See the related notions of 'frame' by Minsky **[145]** and 'script' by van Dijk **[211]**. For a more explicit connection with planning, see Schank and Abelson **[185]**.

the current discourse model or constructing a new one. Integrity constraints regiment such updates. They can take the form of either obligations '?φ `succeeds`', requesting an update of the discourse model satisfying φ, or prohibitions '?φ `fails`', blocking updates of the discourse model satisfying φ. While scenarios describe the meaning of VPs in a tenseless fashion, integrity constraints specify the contribution of tense, aspect and temporal adjuncts to the semantics of VPs.

The third notion is that of *minimal model*. The axioms of the Event Calculus constitute a correct theory of causality if and only if the following two conditions are satisfied:

1. The discourse model contains only those occurrences of events forced to be there by the discourse and the axioms;
2. The interpretation of the primitive predicates is as small as is consistent with the discourse and the axioms.

These two requirements define minimal models. They imply that no unforeseen events are allowed to happen, and that all causal influences are as expected. The choice to work with minimal models instead of all models leads to *non-monotonicity* in discourse interpretation: adding a new sentence or clause to the discourse may invalidate a conclusion derived from the initial minimal model. As shown by van Lambalgen and Hamm [**212**], it is precisely the possibility to retract previously inferred conclusions which allows a rigorous treatment of the imperfective paradox. The most important meta-theorem concerning the EC formalism is that minimal models exist and can be computed efficiently [**212, 194, 196**]. Furthermore, minimal models can be computed (or approximated, depending on the expressiveness of the logical language) by connectionist networks – a topic to which we return in section 4.3.

2.2. The semantics of the progressive. The semantics of the VP 'was writing a letter' can be decomposed into the tenseless lexical meaning of 'write a letter' and the temporal and aspectual contribution of the past progressive. The lexical meaning is represented by the following scenario:[4]

(11) a. $\mathit{Initially}(\mathit{letter}(a))$
b. $\mathit{Initiates}(\mathit{start},\mathit{write},t)$
c. $\mathit{Initiates}(\mathit{finish},\mathit{letter}(c),t)$
d. $\mathit{Terminates}(\mathit{finish},\mathit{write},t)$
e. $\mathit{HoldsAt}(\mathit{write},t) \wedge \mathit{HoldsAt}(\mathit{letter}(c),t) \rightarrow \mathit{Happens}(\mathit{finish},t)$
f. $\mathit{Releases}(\mathit{start},\mathit{letter}(x),t)$
g. $\mathit{HoldsAt}(\mathit{letter}(x),t) \rightarrow \mathit{Trajectory}(\mathit{write},t,\ \mathit{letter}(x+g(d)),d)$

Here *write* is the activity fluent, $\mathit{letter}(x)$ is a parametrized fluent for the completion stage x of the letter, a is a constant for the stage at which writing is initiated, c is a constant for the stage at which the letter is considered finished, and $\mathit{letter}(c)$ is the fluent for the goal state. The dynamics of the scenario is formalized by (11g), which says that, if the stage of completion of the letter at time t is x, then the writing activity, lasting from time t until $t+d$, will result in a letter whose stage of completion at time $t+d$ is $x+g(d)$, where g is a monotone increasing time-dependent real-valued function relating the activity to the completion stage.[5] Accomplishments are distinguished from activities by the statements regimenting the behavior of the goal state, here (11c) and (11e): because no canonical goal is involved in activities, such as 'write letters', their scenario will not contain these clauses.

The contribution of the past progressive is represented by the integrity constraint

(12) $?\mathit{HoldsAt}(\mathit{write},R) \wedge R < S$ `succeeds`

which forces an update of the discourse model such that the activity fluent *write* holds at the reference time R located in the past of the moment of speech S [**175**]. As we noted about

[4]See Andrade-Lotero [**2**] for the algorithms deriving scenarios from discourse representation structures in DRT [**107**] and from constructions in Construction Grammar [**66**].

[5]The way g is defined depends upon the particular VP (or plan) at issue. For instance, 'erasing the blackboard' or 'dismantling the engine' may be represented using a monotone *decreasing* function.

examples (4)-(7), different inferences can be drawn from activities and accomplishments in the past progressive. In the Event Calculus, this follows from the fact that updating the initial model according to (12) leads to different models depending on whether the scenario represents an activity or an accomplishment. The statement below provides information on the inferences licensed by discourses containing an accomplishment in the past progressive.

> COMPLETION INFERENCE *Let $\mathcal{D}$ be a discourse consisting of scenario* (11). *Suppose $\mathcal{D}$ is extended to $\mathcal{D}'$ so that the query ?HoldsAt*($write$, R) $\wedge$ $R < S$ *succeeds in $\mathcal{D}'$. Suppose also* $\lim_{t\to\infty} g(t) \geq c$. *Then there is a unique minimal model of $\mathcal{D}'$ and in this model there is a time $t \geq R$ for which HoldsAt*($letter(c)$, t)*. By virtue of the stipulation Letter*($letter(c)$, t)*, there will be a letter at t.*[6]

If *write* holds at R, as required by (12), it must either hold initially or have been initiated. The latter requires an event *start* which initiated the writing activity. Since the starting event is not provided by discourse, we must assume that *write* holds initially. The clause (11g) states that the writing activity will increase the stage of completion of the letter. As time tends to infinity, the latter will be at least equal to c (the final completion stage). We have stipulated that a letter whose stage of completion is c is a finished letter. Hence, there must be a time at which the letter is considered finished. The writing activity will stop as soon as a complete letter is obtained, as is guaranteed by (11d-e). Notice that this holds for accomplishments, but not for activities: if the VP is the activity 'writing letters', (11c) and (11e) are not part of the scenario and writing will continue also after one letter has been completed – in fact, it will continue indefinitely, if no terminating event is introduced.

We have just seen how the Event Calculus deals with the interpretation of VPs in the past progressive. Now we have to discuss the contribution of subordinate 'when' clauses to the meaning of (4)-(7). Because spilling coffee on the tablecloth usually does not terminate the writing activity, the dynamics of the scenario will lead to a complete letter in (4) and to an indefinite number of letters in (6). The situation is different for examples (5) and (7). Spilling coffee on the paper is typically sufficient for terminating the writing activity. This bit of world knowledge can be expressed by the following addition to scenario (11):

(13) $Terminates(spill, write, t)$

But what is more important here is the integrity constraint introduced by the 'when' clause

(14) $?Happens(spill, R) \wedge R < S$ `succeeds`

which requires the accident to occur at R, while the writing activity was taking place. Since during the writing process there is no complete letter (yet), spilling coffee on the paper will terminate the activity *before* the letter is finished. Therefore, there will be no complete letter in the final discourse structure. As for (7), the theory is consistent both with the situation in which the writing activity was terminated before even a single letter had been completed and with the case in which one or more letters were finished when the accident happened. The proposed semantics predicts that readers would assent to 'The girl has written a letter' in (4), to 'The girl has written no letter' in (5) and to 'The girl has written one or more letters' in (6). This is consistent with received semantic wisdom about entailments of activities and accomplishments in the progressive. Data supporting this view are presented below.

To summarize, the attainment of the consequent state is derived in a minimal model of a discourse containing an accomplishment in the progressive. However, the computation of discourse models is non-monotonic: if the minimal model is extended with a sentence or a clause describing an event which terminates the relevant activity (what we have called a disabling condition), the derivation of the goal state is blocked. Non-monotonicity affords a neat solution of the imperfective paradox, because there is no contradiction in assuming that the representation of the goal state is both essential to the meaning of the progressive VP (as an event type in the scenario) and suppressible on the basis of additional discourse information (as an event token in the minimal model).

---

[6] See van Lambalgen and Hamm [**212**] for a more rigorous statement and a computational proof.

3. An entailment questionnaire

In the preceding sections, we have argued that accomplishments and activities in the past progressive behave differently as regards entailment. An accomplishment such as 'writing a letter' implies that 'The girl has written a letter', provided that no obstacles are described in discourse. Failing to mention a disabling condition is sufficient for concluding that there is no such obstacle to attaining the goal. If however a disabling condition is introduced, the accomplishment will imply that 'The girl has written no letter'. An activity such as 'writing letters' implies that 'The girl has written one or more letters' regardless of further discourse information. Since disabling conditions affect the possibility of attaining a predefined goal, and such predefined goals are part of the meaning of accomplishments but not of activities, accomplishments will be sensitive to the presence of a disabling condition in the discourse context, whereas activities will not. We administered an entailment questionnaire aimed at testing this claim.

3.1. Subjects. Thirty two Dutch native speakers (mean age 22.5, 27 female) completed the cloze probability test (see below) and thirty six (mean age 22.5, 28 female) the entailment questionnaire. Participants were selected from the database of the Donders Centre for Cognitive Neuroimaging in Nijmegen. The two sets of subjects were disjoint.

3.2. Materials. The set of Dutch materials used in the experiments included 160 items. Each item comprised two context sentences (O) providing a neutral setting for the events described by target sentences,[7] four target sentences (A)-(D) and two probe pairs (E)-(F):

(O) De deur van de woonkamer was gesloten. Binnen speelde de radio klassieke muziek.
The door of the living-room was$_{\text{PST}}$ closed$_{\text{PRT}}$. Inside played$_{\text{PST}}$ the radio classical music.
'The door of the living room was closed$_{\text{PST PRT}}$. Inside the radio played$_{\text{PST}}$ classical music.'

(A) Het meisje was brieven aan het schrijven toen haar vriendin koffie op het tafelkleed morste.
The girl was$_{\text{PST}}$ letters on the to-write$_{\text{INF}}$ when her friend coffee on the tablecloth spilled$_{\text{PST}}$.
'The girl was writing$_{\text{PST PRG}}$ letters when her friend spilled$_{\text{PST}}$ coffee on the tablecloth.'

(B) Het meisje was brieven aan het schrijven toen haar vriendin koffie op het papier morste.
The girl was$_{\text{PST}}$ letters on the to-write$_{\text{INF}}$ when her friend coffee on the paper spilled$_{\text{PST}}$.
'The girl was writing$_{\text{PST PRG}}$ letters when her friend spilled$_{\text{PST}}$ coffee on the paper.'

(C) Het meisje was een brief aan het schrijven toen haar vriendin koffie op het tafelkleed morste.
The girl was$_{\text{PST}}$ a letter on the to-write$_{\text{INF}}$ when her friend coffee on the tablecloth spilled$_{\text{PST}}$.
'The girl was writing$_{\text{PST PRG}}$ a letter when her friend spilled$_{\text{PST}}$ coffee on the tablecloth.'

(D) Het meisje was een brief aan het schrijven toen haar vriendin koffie op het papier morste.
The girl was$_{\text{PST}}$ a letter on the to-write$_{\text{INF}}$ when her friend coffee on the paper spilled$_{\text{PST}}$.
'The girl was writing$_{\text{PST PRG}}$ a letter when her friend spilled$_{\text{PST}}$ coffee on the paper.'

(E) Het meisje heeft een of meer brieven geschreven.
The girl has$_{\text{PRS}}$ one or more letters written$_{\text{PRT}}$.
'The girl has written$_{\text{PRS PRF}}$ one or more letters.'
Het meisje heeft geen brief geschreven.
The girl has$_{\text{PRS}}$ no letter written$_{\text{PRT}}$.
'The girl has written$_{\text{PRS PRF}}$ no letter.'

(F) Het meisje heeft een brief geschreven.
The girl has$_{\text{PRS}}$ a letter written$_{\text{PRT}}$.
'The girl has written$_{\text{PRS PRF}}$ a letter.'
Het meisje heeft geen brief geschreven.
The girl has$_{\text{PRS}}$ no letter written$_{\text{PRT}}$.
'The girl has written$_{\text{PRS PRF}}$ no letter.'

Target sentences were constructed manipulating the aspectual class of the VP in the past progressive (activity or accomplishment) and the causal type of the condition introduced by the subordinate 'when' clause (neutral or disabling with respect to the event described in the progressive clause). Activities and accomplishments differed only in the object NP: an indefinite ('een brief') was used for accomplishments, a bare plural ('brieven') was used for

[7] That is, no disabling condition for the event described by the progressive VP was introduced by context items.

activities. Disabling and neutral conditions differed only in the prepositional or object NP, for example 'papier' for the former and 'tafelkleed' for the latter. The distinction between neutral and disabling conditions was made according to the authors' judgment. The probes (E) were used with activities, (F) with accomplishments.

The following linguistic properties of target sentences were normed. The mean length and the raw frequency of the differing words in the NP of subordinate 'when' clauses were matched using the CELEX corpus [**4**]. To determine the cloze probabilities of the verbs in the subordinate clauses, context items followed by a target sentence with the final word blanked were presented to the subjects. Participants were asked to fill in the blank with the first word that came to their mind. Four versions (40 items per condition) randomized and balanced across conditions were constructed. Mean cloze probabilities were not different across conditions (T-tests, all comparisons $P>0.05$). This was established for each version as well as for the entire set of experimental items. Therefore, the same materials and test versions were used in the entailment questionnaire.

3.3. Procedure. Copies of the questionnaire were given to all subjects in the database meeting the following criteria: they had to be native speakers of Dutch with an age between 18 and 50 and with no history of neurological, psychiatric or cognitive disorders. The first 36 subjects who returned a completed questionnaire where included in the analysis.

The first page of the booklet contained the test instructions. Participants were informed that the questionnaire consisted of 160 short texts and that each comprised three sentences (the two context sentences and the target sentence) and was followed by a pair of probes. Subjects were instructed to read each sentence carefully and select the probe which they deemed correct on the basis of their expectations ('verwachtingen') about the continuation of the text ('over het vervolg van de tekst'). Participants were asked to respond as quickly and accurately as possible and to write a brief comment on their answer in the blank space following the probes.

The reference to 'expectations' in the test instructions requires some explanation. Note that we are not interested in what 'subjects semantically know' given a progressive clause – presumably that the writing activity was taking place some time in the past and that only a part of the letter was then completed. For this would amount to asking subjects *what is true at the reference time*, which is captured by the integrity constraint (12) above and requires no inference based on the relevant models. Rather, we are interested in what subjects infer about the outcomes of an action described using the progressive. That is, we want to know how subjects reason about goals and, more precisely, *what is the projected course of events after the reference time*. In order to achieve this, we constructed the probe pairs (E) and (F) using the Dutch present perfect (which focuses on the present consequences of a past event), we avoided probes of the form 'The girl has written a part of the letter', and we asked subjects to decide which of the two probes (positive or negative) matched their 'expectations' about the continuation of the narrative. Participants' written explanations provide no evidence that (E) and (F) were insufficient for giving accurate responses, such that, for instance, a probe of the form 'It is unclear whether the girl has written a letter' was necessary. Rather, subjects' comments suggest that, if the discourse implied that only a part of the letter was completed, as in (D), then the negative probe in (F) had to be selected.

3.4. Data analysis. Subject-based and item-based statistical analyses were carried out. For the former, we used a repeated-measures ANOVA model with Subject as the random effect, Aspectual Class (Activity/Accomplishment) and Condition Type (Neutral/Disabling) as fixed effects, and the mean number of negative responses as the dependent variable. In order to generalize over both subjects and test items, we used a parallel repeated-measures ANOVA model in which Test Item (as defined above, that is, as a set of context, target and probe sentences) was the random effect, Aspectual Class and Condition Type were the fixed effects and the mean number of negative respondents (subjects giving a negative response) was the dependent variable. Univariate F-tests were computed in both cases.

	Subject-based analysis	Item-based analysis
Aspectual Class	$F(1,35)=64.763$ $P<0.001$	$F(1,159)=619.240$ $P<0.001$
Condition Type	$F(1,35)=112.560$ $P<0.001$	$F(1,159)=237.270$ $P<0.001$
Aspectual Class × Condition Type	$F(1,35)=61.832$ $P<0.001$	$F(1,159)=100.210$ $P<0.001$

TABLE 1. Summary of ANOVA statistics for the questionnaire data.

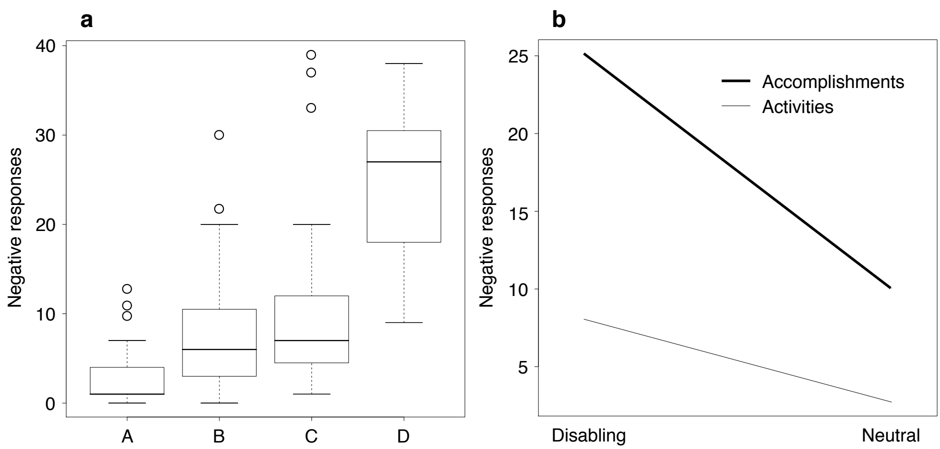


FIGURE 1. Results of the questionnaire. (a) Boxplot for the subject-based analysis. Solid lines within the boxes represent the median, box height is equal to the interquartile range, whiskers indicate adjacent values, empty circles denote outliers. The maximum of potential negative judgments is 40. (b) Interaction plot for the subject-based analysis. Traces endpoints represent the mean number of negative responses in each experimental condition. The maximum of potential negative judgments is 40.

3.5. Results. The subject-based analysis revealed significant effects of Aspectual Class and Condition Type and a significant interaction between the two factors (table 1). Neutral activities (A) had the lowest mean of negative responses (M=2.72, SD=3.22), followed by disabled activities (B) (M=8.06, SD=7.05), neutral accomplishments (C) (M=10.03, SD=9.23) and disabled accomplishments (D) (M=25.14, SD=8.02). Except for neutral activities (A), the distribution of the data in the different conditions appears rather similar, as indicated by standard deviations (SDs) as well as box height and whisker length in figure 1-a. Figure 1-b shows that the pattern of responses is largely as predicted by the theory: accomplishments are more sensitive than activities to the presence of a disabling condition in the discourse context. A similar pattern of effects was revealed by the item-based analysis (see table 1). Neutral activities (A) had the lowest mean of negative respondents (M=2.35, SD=4.13), followed by disabled activities (B) (M=7.13, SD=7.24), neutral accomplishments (C) (M=9.54, SD=7.09) and disabled accomplishments (D) (M=22.62, SD=9.04).

4. Immediacy and non-monotonicity

In the preceding sections we have proposed an analysis of the imperfective paradox based on the Event Calculus and we have presented some behavioral data supporting the theory. Recall that we chose to design the semantic theory as a computational analysis in Marr's sense. We are now ready to move on to the intermediate level of analysis, and consider the processing steps subserving the computation of discourse models.

The processing model is based on the combination of Event Calculus and Constraint Logic Programming by van Lambalgen and Hamm [**212**]. The algorithms that are presented there specify in an abstract manner the computational steps involved in satisfying integrity constraints or, more precisely, they spell out the derivation (the proof) of a given statement within the theory. In order to derive predictions about the complexity and the time-course of the relevant computations, it is necessary to add an explicit processing component to the formal machinery of the Event Calculus: the principle of immediacy.

4.1. The principle of immediacy. In a paper discussing language processing models and their neural realizations, Hagoort [**74**] proposed "six general architectural principles for comprehension beyond the word level". One of these, also known as the 'principle of immediacy', is particularly relevant in this context. Immediacy has often been debated in the psycholinguistic literature [**133, 99**], where it is viewed as a hypothesis about the time-course of the access and integration of lexical meanings in a sentence structure and is often contrasted with syntax-first models [**58, 61**]. Immediacy is defined by Hagoort [**74**] as

> the general processing principle of unification. Semantic unification does not wait until all relevant syntactic information (such as word class information) is available, but starts immediately with what it derives on the basis of the bottom-up input and the left context. The corollary of immediacy is incrementality: output representations are built up from left to right in close temporal contiguity to the input signal.

This statement cannot be directly related to our Event Calculus analysis, the reason being that it covers only lexical integration ('unification') and not the computation of denotations and discourse models. Hagoort's notion of immediacy is sufficient for deriving processing predictions concerning semantic composition. However, to be relevant for describing the construction of discourse models, the principle of immediacy must be thus reformulated:

> IMMEDIACY *The algorithms updating a minimal model so as to satisfy an integrity constraint are executed as soon as the integrity constraint is given as input.*

The proposed definition of immediacy is a general hypothesis about the time-course of the construction of discourse representations. It implies that minimal models are computed as soon as update instructions, in the form of integrity constraints, are fed into the system by means of the relevant lexical and morpho-syntactic material.

4.2. The recomputation hypothesis. Let us now consider the processing steps leading to the construction of discourse models for (C) and (D). The first update instruction, in the form of integrity constraint (12), is introduced by the progressive VP. Since at that stage the information provided by discourse amounts to the context (O) and the main (progressive) clause of (C) or (D), it is assumed there is no obstacle for the activity. Immediacy implies that (12) will be satisfied as soon as the progressive VP is processed, and the Event Calculus guarantees that a minimal model is computed: there is a time at which the consequent state holds. The prediction is that, when an accomplishment in the past progressive is processed, a minimal model is immediately computed in which the goal state is attained.

Considering the subordinate clause, we can distinguish two processing steps. The first is carried out by adding (13) to the scenario (11), updating the discourse model according to (14), and deriving the existence of a time later than the initiation of the writing activity at which writing was brought to an end. The second step computes the suppression of the goal state inference by deriving the further statement that there is no time at which a complete letter was obtained. These are two distinct operations: for activities that are terminated by a disabling condition, such as (B), only the first step is carried out when the subordinate is processed because there is no goal state to be canceled; for disabled accomplishments, such as (D), both steps are implemented. Recall that the satisfaction of the goal state is derived in a minimal model of the progressive VP. As a consequence, the hypothesis licensed by the theory is that, when an accomplishment in the progressive is followed by a subordinate clause describing a disabling condition, the initial minimal model is *recomputed* to the effect that, in the new discourse structure, there is no time at which the goal state holds.

Recomputation is thus the main processing consequence of the imperfective paradox. Predictions concerning the complexity and the time-course of semantic computations for sentences like (A)-(D) can be derived from our processing model. The theory implies that the subordinate clause in (D) involves the recomputation of the minimal model computed while processing the progressive clause, while in (C) the initial model is simply extended (we return on the difference between 'recomputation' and 'extension' in 4.3). Furthermore, the activity cases (A) and (B) will also involve a monotonic extension of the initial model, such that the termination of the writing activity is computed for (B), but no cancelation of the goal state, because no canonical goal is involved in activities. In short, the subordinate clause in (D) might be more complex to process compared to (A)-(C), as no recomputation is triggered in the latter conditions. As regards time-courses, the Event Calculus requires the causal and temporal information carried by the verb 'morste' to activate the additional scenario clause (13), satisfy (14), and derive the failed occurrence of the consequent state. Therefore, recomputation is expected to surface only at the final word in (D).

4.3. Recomputation and perceptron learning. The reader may ask why computing a model that is incompatible with a previous structure (recomputation) would be different from, and in particular more complex than, computing a minimal model that monotonically extends the initial one. The answer can be found in the behavior of the neural networks that compute minimal models. Due to the syntactic restrictions inherent in Logic Programming [**39**], the models characterized by the Event Calculus can be viewed as the stable states of the associated neural networks. Recurrent neural networks can compute minimal models for propositional logic programs [**194, 196**], and for any propositional logic program there exists a 3-layer feedforward network of binary threshold units that computes the semantic operators on which the construction of minimal models is based [**91**]. The language of the Event Calculus is not propositional but is a many-sorted predicate logic (see section 2.1), with matters being further complicated by the use of integrity constraints. Recent research suggests however that recurrent networks can also approximate the semantic operators for first-order logic programs and their fixed points [**91**]. Furthermore, integrity constraints can be modeled via a form of back-propagation called 'perceptron learning' [**196, 179, 178**].

In this framework, any computation on a given minimal model, such as adding a logic program clause (a scenario clause), will somehow bring the network from its initial stable state to another stable state, corresponding to the updated minimal model. Nonetheless, if the neural representation proposed by Stenning and van Lambalgen [**196**] is approximately correct, there is a large difference in neural activity between, on the one hand, a monotonic extension of a minimal model and a non-monotonic recomputation of a minimal model on the other. Consider first the case of a monotonic extension as envisaged by our processing model. Retrieving a clause such as (13) from semantic memory, assuming that *spill* denotes 'spilling coffee on the *tablecloth*', will result in the activation of a number of units ('neurons') which were previously silent; but the activation state of the remaining units, including those representing the goal state (the complete letter), will remain the same. However, retrieving a different clause from semantic memory, for instance (13) where *spill* now denotes 'spilling coffee on the *paper*', will result in the activation of neurons which were silent *but also* in the readjustment of the activation patterns of units which were previously active. For instance, the neurons representing the consequent state (the complete letter) will no longer be active. This change in activation is achieved in the neural network by applications of perceptron learning. The difference between monotonic extension and non-monotonic recomputation can therefore be found in the occurrence of the readjustment of connection strengths driven by perceptron learning. Efforts towards a more realistic neural implementation of minimal models and non-monotonic computation will be discussed in later chapters.

These considerations imply that non-monotonic recomputation in (D) has more drastic consequences on neural processing as compared to the monotonic extension of a minimal model in (C). We must now show that a processing model based on monotonic semantics would not predict a similar effect. In a strictly monotonic progression of structures, the goal state is necessarily not represented, for otherwise the model may have to be recomputed, which is not allowed by the monotonic logic underlying the theory. In the Event Calculus framework, this means that the predicates and axioms for continuous change are not used, at most those for instantaneous change. In addition, because the chain of discourse models is monotonic, one never actually computes minimal models.[8] So, under the assumption of monotonicity, little semantic computation is going on. In particular, (C) and (D) will both lead to monotonic extensions of the initial model, and both extensions will be computed at about the same time, that is, when the final verb is encountered: in (C) the update will give rise to a model in which (given the results of our entailment questionnaire) the consequent state is represented as being attained, in (D) it will lead to a model in which the consequent state does not hold. Because there is no principled semantic or processing reason to assume that one final model would be more complex to compute than the other, it seems to follow that a strictly monotonic semantics predicts no difference between the conditions.

The recomputation hypothesis can be tested in a dedicated electrophysiological study. An ERP experiment using the materials of our questionnaire as stimuli would allow us to determine whether recomputation occurs or, more precisely, whether there is a processing difference between (D) and (C) which cannot be accounted for in terms of lexically-bound factors such as frequency, cloze probabilities, ambiguity etc. Preliminary ERP evidence for recomputation is presented in chapter 5.

[8]Formally, one never computes the completion of the scenario plus the EC axioms. For the notion of 'completion', see **[212]**.

CHAPTER 3

The processing consequences of compositionality

This chapter is a modified version of G. Baggio, M. van Lambalgen & P. Hagoort. The processing consequences of compositionality. In W. Hinzen, E. Machery & M. Werning (Eds.), *The Oxford Handbook of Compositionality*. Oxford University Press, to appear.

1. Introduction

It is often said that the principle of compositionality is formally vacuous: any grammar can be given a compositional semantics [**102, 224, 109, 222**], which implies the principle is also empirically vacuous: if a compositional treatment of any linguistic structure can be given, then compositionality is always upheld by the data. To be sure, the meaning of any complex expression can be viewed as a function of the meanings of its constituents and the syntactic mode of combination, provided enough complexity is built into the structures involved, i.e. syntax and the lexicon. These need not be motivated on independent grounds, since their characterization serves the sole purpose of yielding a compositional theory [**69**].

The need for an independent motivation of theories of syntax and lexical semantics is precisely the issue here. Our aim is to show that, even though there often is a way to salvage compositionality in the face of empirical data, the choices one has to make in order to do so have consequences which may be implausible given the cognitive and neural constraints on language comprehension, production and acquisition. Let us start with the most basic of questions: *why* compositionality? We will now give a sketch of the main arguments, which will be refined in the course of the discussion.

1.1. The productivity argument. One familiar argument in favor of compositionality starts from the perceived tension between the infinity of language and the finiteness of the brain. There are infinitely many sentences in any natural language, but the brain has only finite storage capacity, and it therefore falls to syntax to give a finitely describable procedure for generating an infinite class of sentences. Furthermore, so the argument goes, a speaker of any language is able to understand a sentence she has never heard before, or to express a meaning she has never expressed before, and in that sense she knows the infinitely many different meanings of the infinitely many sentences of that language. Therefore, semantics is also under the obligation to come up with a finitely describable engine that generates all possible sentence meanings for the given language [**108**].

Compositionality provides a seemingly efficient way to satisfy these desiderata. There are only finitely many words in the lexicon and syntax can have only finitely many rules of combination. Here compositionality comes into play:

> PRINCIPLE OF COMPOSITIONALITY *The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined.*

If meanings are generated in such a way that compositionality is satisfied, then it seems that all possible sentence meanings can be finitely generated. Now, although compositionality is a guiding principle of *formal* semantics, the standard motivation as sketched above partly appeals to such *psychological* notions as comprehension and production, while at the same time invoking the patently non-psychological infinity of language. A quick way to dismiss the argument from productivity is therefore to deny that language is infinite, even in the sense of *potential* infinity. A moment's reflection shows however that the issue is not really about infinity – substituting a large finite number for infinity does not change the essence of the productivity argument, which is that not every sentence that can be understood or produced given human cognitive limitations is stored. So, while there is no reason to have a semantic theory that explains comprehension of nested center embeddings of arbitrary depth, it is also not the case that all sentences with center embeddings with depth, say, ≤ 6 can be stored. In other words, psychologically speaking, the real issue is *the balance between storage and computation*, and the role compositionality should play there. It might seem that compositionality always leads to the most efficient architecture in this respect.

That this is not necessarily so, can be illustrated using an example from Keenan [**110**]. In an adjective–noun construction, the noun is the argument fed into the adjective, which is viewed as a function. Keenan observes that the interpretation of the function word seems to be determined by its argument: compare for instance the different meanings of the adjective 'flat' in 'flat tyre', 'flat beer', 'flat note' etc. It is of course technically possible, as Keenan notes, to replace the single function 'flat' by a disjunctively defined function, where each of the disjuncts corresponds to a separate meaning for 'flat', with suitable selection restrictions on the argument. However, this technical solution is surely paradoxical. Compositionality was invoked to account for productivity, which seemed hard to explain in terms of storage only. But, in this case, compositionality can apparently be salvaged only by *increasing* the demand on storage. From a processing perspective, it would be much better if there were a single computational mechanism generating the meaning of a *flat + noun* construction, starting from a single basic meaning of 'flat'. These considerations show that the principle of compositionality is affected by its ambiguous status: as a formal desideratum on the one hand, and a processing hypothesis on the other.

1.2. The systematicity argument. A second argument in favor of compositionality is based on the observation that languages are systematic, that is, the ability to understand certain utterances is connected to the ability to understand certain others. For instance, any native speaker of English that understands 'John loves Mary' understands also 'Mary loves John'.

> SYSTEMATICITY *Anyone who understands the complex expressions e and e' built up through the syntactic operation F from the constituents $e_1, \ldots, e_n$ and $e'_1, \ldots, e'_n$ respectively, can thereby also understand any other expression e'' built up through F from expressions among $e_1, \ldots, e_n, e'_1, \ldots, e'_n$.*

Systematicity seems to entail compositionality, but the issue here is whether languages are systematic in the technical sense. For instance, anyone who understands 'The dog is asleep' and 'The cat is awake' can certainly understand 'The dog is awake' and 'The cat is asleep'. However, would everyone who understands 'within an hour' and 'without a watch' also understand 'within a watch' and 'without an hour'? The definition presupposes that e'' is a meaningful expression, and that $e_1, \ldots, e_n, e'_1, \ldots, e'_n$ can be freely combined and substituted while keeping F constant. But the fact that we can hardly make sense of 'within a watch' and 'without an hour' suggests that this is not the case, thus languages are not systematic in the sense of the definition above. Let us suppose for a moment that this difficulty can be overcome. Would then systematicity force compositionality upon us? It seems it wouldn't, for systematicity says that, given the sentences 'The dog is asleep' and 'The cat is awake', from the meanings of 'the dog', 'the cat', 'is asleep' and 'is awake' plus the syntax, one is able to understand the meaning of 'The dog is awake'. Compositionality makes however a stronger claim, namely that the meanings of 'the dog' and 'is awake' plus the syntax are sufficient to do that. Thus, even if systematicity held, it would not buy us compositionality (see Johnson [**104**], Szabó[**198**], and Pullum and Scholz [**166**] for a discussion).

1.3. The methodological argument. A third argument assumes that compositionality is needed as a constraint on doing semantics, that is, as an essential part of the explanatory enterprise [**103, 37**]. For instance, if one has to explain why in the 'donkey sentence'

(1) If a farmer owns a donkey, he beats it.

the DP 'a donkey' has universal force, it won't do to say: "well, in this context it simply has universal force". An account that starts out with the existential reading of the DP, and then shows how its being embedded in the antecedent of a conditional changes its interpretation from existential to universal, has at least the appearance of an explanation.

The trouble with the methodological argument is that compositionality is highly theory-dependent [**159**]. Ideally, when looking for an explanation of a linguistic phenomenon, one assumes the syntax and the semantics to be fully specified formal systems. In that case, it is a definite question whether that phenomenon allows for a compositional treatment, and if so, to what extent. If it does not, one may take this as a cue for changing the semantics. In practice, however, the explanation of a new phenomenon of interest often leads to changes in *both* syntax and semantics. Compositionality becomes a soft constraint indeed.

The much-needed methodological constraints have to be sought elsewhere, in a tighter regimentation of syntactic and semantic theories. From our perspective, these constraints should be cognitively motivated, in the sense that formal theories of syntax and semantics should be viewed as 'computational level theories' in Marr's [**131**] sense of actual syntactic and semantic mechanisms. In the ideal case, it is then an empirical question whether syntax and semantics communicate as compositionality says they do.

1.4. The modularity argument. A fourth argument for compositionality is suggested by a notion of the language faculty as a 'cognitive module'. Fodor [**55**] lists nine properties characterizing modular systems: domain specificity, fast and mandatory operation, limited central access to representations in a module, informational encapsulation, shallow outputs, fixed neural architecture, specific breakdown patters, and characteristic ontogenic pace and sequencing. Of these, the most relevant for our purposes is *informational encapsulation*. This is the idea that perceptual systems – language included – are relatively impenetrable to the bulk of the knowledge that is internally represented. Informational encapsulation says that there are tight constraints on the flow and the handling of extra-modular information *within* a module *prior to* the production of an output.

To a certain extent, informational encapsulation is assumed by any cognitive model of language – which is not to say that all component-based architectures [**98**] are modular in Fodor's sense. Fodor's – and Chomsky's – original view of modularity is that a grammar's generative power can be captured by a *single* module, which comprises a finite repository of lexical meanings and a finite repertoire of syntactic rules.[1] Rules for semantic computation (inference is a paradigmatic case) would then fall within the province of central systems. It can be easily seen that the computations performed by this kind of modular machine are those regimented by compositionality: the output produced by the module (the meaning of a complex expression) is a function of the knowledge accessible to the module (the lexicon and the syntax). But this is not the only modular architecture supporting compositionality. For instance, one could postulate *two* modules: a module which produces syntactic analyses of clauses, which are then fed into a second module that contains meanings for the lexical items and combination procedures that mirror syntactic operations, and outputs a semantic representation of the clause.[2] Compositionality would then constrain the kind of traffic that can occur between the two modules.

Regardless of the choice of modular architecture, compositionality seems to be relevant insofar as it acts as a counterpart of information encapsulation at the level of description of linguistic form. The connection between compositionality and informational encapsulation can be made more explicit: if the composition of meanings is not affected by extra-modular knowledge, then one can characterize the meaning of any complex expression as a function of the meanings of its constituents and syntactic rules *only*, all of which are readily available

[1] The module also contains mechanisms for phonological decoding, but we ignore these for simplicity.

[2] For a modular view of semantics, see among others Borg [**10**] and Robbins [**177**] for a critical discussion.

within the module.[3] Clearly, this hinges very much on what one assumes is contained in the module(s) – and this will be a recurrent theme in this chapter. What bears some emphasis here is that the degree to which a system is informationally encapsulated can be determined only based on empirical data. Hence, with this 'argument from modularity' in place, it also becomes possible to treat compositionality as a processing principle, that is, as a constraint on the range of data structures involved in language processing.

2. Compositionality as a processing principle

2.1. A first approximation. How can one constrain and refine compositionality based on experimental data and cognitive considerations? One may start from the observation that 'function' in the definition of compositionality needs to refer to some computable input-to-output mapping, and that the input – lexical meanings and syntactic constraints or rules – must be given incrementally:

> INCREMENTAL COMPOSITION *The meaning of a complex expression at some processing stage* σ *is computed based on the constituent expressions processed at* σ *and the syntactic structure built up at* σ.[4]

This definition is silent as to whether meaning assembly involves the lexicon and the syntax *only*, or whether other sources of information could enter the composition process. For this we need another definition, which can be combined with the former:

> SIMPLE COMPOSITION *(i) The meanings of elementary expressions are the only constraints on content in the computation of the meaning of a complex expression. (ii) The syntax of elementary expressions is the only constraint on structure in the computation of the meaning of a complex expression.*

The notion of 'weak' or 'enriched composition' [**171, 98**] follows from the choice to allow further constraints on content and structure. The distinction between content and structure is admittedly vague, and it can only be made clearer based on particular formal theories of syntax and semantics: recall compositionality's theory-dependence [**159**]. We could go even further and observe that, at least in formal approaches to grammar, the distinction vanishes, and one is left with a purely syntactic analysis of meaning in some logical language. Still, it remains a distinction worth keeping, especially because the brain appears to honor it, as is reflected by the different electrophysiological traces left by morphosyntactic and semantic manipulations [**117, 151, 78**]. Let us now see how two of the arguments for compositionality presented above fare in light of the new definitions.

2.2. Productivity and modularity reconsidered. The plausibility of compositionality is usually argued for by means of a rhetorical question, like "what other mechanism could explain the productivity of language?", as if posing the question would already dictate the answer. To address this point, it pays to be more precise regarding the technical implications of simple composition. Consider two supposed consequences, due to Hintikka [**89**]:

> CONTEXT INDEPENDENCE THESIS *The meaning of an expression should not depend on the context in which it occurs.*
>
> INSIDE-OUT PRINCIPLE *The proper direction of a semantic analysis is from the inside out; from the bottom up; from elementary to complex meanings.*

Taken together, these notions suggest that we ought to take the concept of a 'function' in the formulation of the principle very seriously. Semantic computation of complex expressions is function application, so *the meanings of simple expressions are not changed by virtue of the fact*

---

[3] Jackendoff [**98**] makes the same point: "The hypothesis of syntactically transparent semantic composition has the virtue of theoretical elegance and constraint. Its effect is to enable researchers to isolate the language capacity – including its contribution to semantics – from the rest of the mind, as befits a modular conception."

[4] For simplicity, we may assume that each word corresponds to a processing stage, and vice versa. An important theoretical question is whether assuming finer-grained processing steps would lead to local inconsistencies between incrementality and compositionality.

that they occur as arguments of the function.[5] This shows that the principle of compositionality is tied to one very particular form of linguistic productivity, exemplified by the usual rules: if S is a sentence, one can form a new sentence 'I think that S'; if S_1, S_2 are sentences, one can form 'S_1 and S_2'; etc. But there exist other forms of productivity in natural languages, which do not have this function-like character. An example is the progressive construction in English when applied to stative verbs [**30**]:

(2) a. She resembles her mother.
b. She is resembling her mother.*
c. She is resembling her mother more and more every day.

'Resemble' is a stative verb, and this seems to be clinched by (2b), which clearly shows the progressive is not applicable. Still, in a suitable context the progressive *is* applicable, as in (2c), where it imposes a dynamic meaning upon 'resemble'. Here resemblance comes in degrees that can change over time. Therefore, the meaning of 'resemble' depends upon the context in which it occurs, contradicting context independence. This variability of meaning can still be predicted once one assumes that the progressive construction has a meaning of its own, which it imposes upon that of the verb. This imposition of meaning is productive in that it applies to several stative verbs. Simple composition as made precise by Hintikka can account for this particular form of productivity only by assuming multiple meanings of stative verbs, where the progressive selects for a dynamic meaning.

While compositionality can be salvaged formally, experiments on language processing may well rule out such *ad hoc* maneuvers. There is a very different computational account of what goes on in comprehending sentences such as (2c) that emphasizes the recomputation of the meaning of 'resemble' which takes place when the adverbial phrase 'more and more every day' is processed [**212**, chapter 11]. The two accounts thus differ in their processing consequences. Simple composition in the Hintikka sense leads to a search and selection of the appropriate meaning though not to recomputation, as the second account instead does. These operations might give rise to different neurophysiological responses, as is suggested by the data reported in chapter 5. The two accounts can be tested experimentally.

If recomputation were to be supported, where would that leave compositionality and the work it is supposed to do? The strict reading embodied in Hintikka's principles presents semantic composition as entirely inside-out/bottom-up. The recomputation account is also partially outside-in/top-down. In theory, this has the consequence that the meaning which is assigned to an expression is always provisional, a situation that cannot occur on a literal reading of compositionality (but see 3.1). However, there is room for both accounts, because the productivity of language is a two-dimensional phenomenon. On the vertical dimension, there is productivity due to increased syntactic complexity: here simple composition has an important role to play. There is however also a horizontal dimension to productivity: here it is not so much the syntactic complexity that increases, but new meanings of a given clause are produced by varying syntactic or semantic aspects of other clauses. Thus, if we replace the adverbial 'more and more every day' in (2c) with 'in broad daylight', the progressive is no longer felicitous, because the verb 'resemble' then gets its default stative meaning. The horizontal dimension of productivity seems to call for some form of enriched composition, that is, ways of combining meaning that allow top-down influences and recomputation.

Two forms of top-down computation are consistent with informational encapsulation, and thereby with its analog compositionality. First, information that is fed back comes from and remains within the module, as when it is stored in the lexicon. Second, information is fed back into the module from another module or from a central system *after* the production of an output; this is a rather trivial way to preserve informational encapsulation in the face of top-down computation. The former can be dismissed using an argument due to Hodges [**92**], based on what we might call 'top-down composition'. It takes its cue from Frege [**59**]:

[5] Just as the number 2 has the same meaning, whether it occurs as argument in the function 3^x or in the function $3+x$. This is obvious in mathematics, but not in natural language, as we shall see.

CONTEXT PRINCIPLE *Elementary expressions do not have meaning in isolation, but only in the context of (as constituents of) complex expressions.*

In briefest outline, Hodges' proposal is this. The syntax is defined in terms of constituent structure in such a way that, if e is an expression occurring in a sentence S, then S can be viewed as $G(e)$, where $G(x)$ is a syntactic frame with open slot x. Now define an equivalence relation $\sim$ on expressions by putting

$e \sim f$ iff for all $G(x)$: $G(e)$ is a sentence iff $G(f)$ is, and $G(e)$ is acceptable in the same contexts as $G(f)$.

The *Fregean value* of e is the equivalence class of e modulo $\sim$. Hodges shows that taking the Fregean value of e as its meaning yields a compositional semantics. Therefore, if we assume that the module contains all Fregean values, modularity is restored. Moreover, this notion of meaning is pleasingly context-sensitive. Hodges gives the example of sentences which have different social connotations, for instance 'he is intoxicated' *vs.* 'he is pissed' **[92]**. The contexts in which these sentences are acceptable are very different. In this sense, these two sentences have different meanings in Hodges' semantics, whereas they would be treated as synonymous with 'he is drunk' in a more standard framework.

Although Hodges' proposal formally restores compositionality, it does so in a way that upsets the balance between storage and computation, and renders it unclear how meanings could be acquired. The Fregean value of e is defined by means of all possible uses of e, and it is doubtful that these are available to a young language learner. In other words, learning meanings is a gradual process, thus Fregean values should be partial objects, in the sense of being defined by means of a subset of all possible uses of e, and being subject to update and revision as learning proceeds – which is not what Hodges suggests. The point can be amplified by considering what it means to *know* the Fregean value of e. There seem to be two components to this:

(i) one must know which f are equivalent to e, which means that for all $G(x)$ such that $G(e)$ is a sentence iff $G(f)$ is, $G(f)$ is acceptable in exactly the same contexts as $G(e)$ – this requires one to know for all sentences $G(e)$ and all contexts C, whether $G(e)$ is acceptable in C;

(ii) one must know which f are not equivalent to e, which means that for all G such that $G(e)$ and $G(f)$ are sentences, one must know a context in which $G(e)$ is acceptable and $G(f)$ is not, or vice versa.

Natural (as opposed to formal) languages may not incorporate the concise representations generating the knowledge required. This implies that the storage component must already contain a great deal of information about the sentences that can be constructed from e, and the contexts in which these are acceptable. Intuitively, this goes against the very purpose of a modular architecture, and it also goes against the original motivation for compositionality as easing the burden on storage – recall the argument from productivity. We therefore tend to read Hodges' result, when applied to natural languages, as showing the implausibility of an architecture in which context sensitivity is achieved by storing extra information within the module, rather than by relaxing informational encapsulation to allow cross-module talk. In brief, the *a priori* arguments considered here show that simple composition is not enough to account for the full range of factors which make language productive. Let us now ask if similar considerations are suggested by experimental work bearing on compositionality.

3. Compositionality and processing data

3.1. Semantic illusions. An assumption that underlies many semantic theories is that only full lexical representations are used in meaning composition. In most proposals, this choice is forced by the ontology: a lexical meaning is a placeholder for a typed entity, which explains why inputs cannot be partial objects. Therefore, meaning assembly reduces to type composition. This accounts for the fact that inputs are still recognizable in the end product. In lexical semantics, by contrast, lexical meanings are complex, and yet compact structures, such as algorithms [**144**], feature or conceptual structures [**96**] etc. One possible refinement of compositionality would then be to allow partial representations to be recruited during processing.

Relevant to this issue is the well-known psychological phenomenon called the 'Moses illusion' [**45**]. When asked 'How many animals of each sort did Moses put on the ark?', subjects tend to respond 'two', without questioning the (false) presupposition that Moses was the biblical character who did that. Similar results have been obtained with questions such as 'After an air-crash, where should the survivors be buried?' [**5**] and 'Can a man marry his widow's sister?' [**183**]. Hearers seem to be processing these sentences superficially enough to miss the false presuppositions, but Ferreira et al. [**49, 50**] go as far as suggesting that these data challenge compositionality. Indeed, the meaning of the Moses question computed by hearers and the meaning derived compositionally are in some important respect different. For if the former were a function of the meanings of the constituents and the syntax, then 'Moses' would mean Moses and hearers would notice the false presupposition. This seems an instance of a non-compositional process.

One might argue against this conclusion by emphasizing that these data just show that not every word in the sentence need contribute its full meaning [**184**]. The 'full meaning' of 'Moses' may not be retrieved, but only some feature made salient by the context [**184, 183**]. This may be 'biblical character', 'patriarch', or another feature responsible for the semantic proximity of 'Noah' and 'Moses' [**215**]. Feature sharing is, on this account, what gives rise to the illusion. Crucially, the fact that the lexicon may be a locus of shallow processing (or retrieval, as the case may be) does not speak against compositionality. Simple composition entails that the lexicon is the *only* provider of content for complex meanings, though not that *full* lexical representations must be used. The latter would be too strong a requirement to force upon either simple or enriched composition. For if there exist such entities as 'full lexical meanings' – and there are reasons to be suspicious about that – they can hardly be used on most processing tasks, because of the massive amount of information, presumably continuous with non-lexical knowledge, that would be fed into the composition process. Incremental, context-sensitive feature selection during word meaning retrieval seems to be a much more plausible default assumption. Semantic illusions can thus be seen as a special case, in which some critical semantic feature is shared between the word triggering a true presupposition ('Noah') and the word triggering a false presupposition ('Moses'). In brief, *pace* Ferreira et al., compositionality can be refined to accommodate semantic illusions, by allowing the composition process to make use of partial lexico-semantic representations.

3.2. Lingering misinterpretations. Ferreira et al. studied cases of misinterpretation in language processing. Some of these involve garden-path sentences:

(3) While Anna dressed the baby played in the crib.

The 'garden-path model' [**58**] hypothesizes that 'the baby' is initially parsed and interpreted as a direct object of 'dressed'. Only when the verb 'played' is encountered are the syntactic and semantic representations revised to the effect that 'the baby' is the subject of 'played'. One issue here is whether the initial representation, featuring 'the baby' as a direct object, is at all maintained in memory. Christianson et al. [**25**] show that, while readers correctly respond for the affirmative to 'Did the baby play in the crib?', they also answer positively to 'Did Anna dress the baby?'. No grammar or parser on the market would allow the same token NP to play two functional roles, subject and direct object. Nonetheless, this appears to be precisely the interpretation arrived at by readers. Ferreira et al. [**49**] take this as "clear evidence that the meaning people obtain for a sentence is often not a reflection of its true content" – that is, it is not built up compositionally.

Does the existence of 'lingering misinterpretations' demonstrate that the processing of garden-path sentences is non-compositional, as suggested by Ferreira and colleagues? There are at least two ways of accounting for the data, from which different answers ensue. On one account, the last interpretation subjects come up with is that while Anna dressed, the baby played in the crib, which corresponds to the revised parse whereby 'the baby' is the subject of 'played', and is no longer the direct object of 'dressed'. Interpretation is therefore non-monotonic, that is, entailing the revision of an earlier structure. This however requires a refinement of compositionality that was introduced earlier on – incremental composition. On this view, both the initial and the revised interpretations can be derived compositionally, and simple composition seems enough in this case. The persisting misinterpretation would be taken as an effect of memory architecture or neural implementation. One aspect of the data of Christianson et al. [**25**] which supports this story is that misinterpretations are more frequent when the head of the misanalyzed phrase occurs early. That is, misinterpretations are more likely to persist the longer they have been part of the initial discourse model.

On the second account, the final interpretation is that Anna dressed the baby while the baby played in the crib. This meaning can hardly be derived compositionally. First, because there is just a single token of 'the baby' among the constituents of the sentence, whereas the final interpretation of the sentence contains two occurrences of it: one as the recipient of the dressing action, the other as the agent of the playing activity – hence the 'constituent parts' aspect of the definition of compositionality is out. Second, because syntax does not allow a phrase to play two distinct functional roles simultaneously – therefore the 'syntax' part of the definition is out. To derive the meaning above, one needs a mechanism that copies the token 'the baby' and makes both instances available for interpretation. Such a mechanism does make processing non-compositional [**49, 50**]. Thus, misinterpretations of garden-path sentences challenge compositionality, unless we assume that early semantic material lingers in memory also during later stages, but is not part of the discourse model computed on the basis of the revised syntactic analysis.

3.3. Event-related brain potentials. ERPs have already been mentioned in chapter 1. Here we will skip technical details as much as possible and focus on the relevance of known ERP components for compositionality. One key aspect of enriched composition is that the stored meanings of elementary expressions are not the only source of content for complex meanings. This challenges informational encapsulation, if one can demonstrate that such additional semantic information is handled by the module *before an output is produced* [**55**]. ERPs allow one to make timing inferences of the kind required to address this issue [**210**].

An ERP component that is particularly relevant here is the N400. This is a negative shift starting around 250 ms after word onset, peaking at 400 ms and lasting until approximately 550 ms. Every content word elicits an N400, but the amplitude of the component, relative to some control condition, is dependent upon the degree of semantic relatedness of the given word with its sentence [**117, 118, 76**] or discourse context [**208, 209**]. There is evidence that the N400 does not merely reflect lexical access, but also the *integration* of a word's meaning into the unfolding semantic representation [**75**]. Tentatively, the N400 can be considered as an index of the complexity of initial attempts to combine the meaning of a given word with the meanings of the expressions already processed.

Another ERP effect that is of interest here is the P600. This is a positive shift starting around 500 ms following the onset of the word and lasting for about 500 ms [**151, 78**]. Larger P600 effects are elicited by sentences containing violations of syntactic constraints (such as phrase structure, subcategorization and agreement), temporarily syntactically ambiguous sentences, garden-path sentences, and constructions which show high syntactic complexity [**77, 61, 71**]. In relation to incremental composition, we can regard the P600 as an index of the time and resources involved in attaching a given word to the syntactic representation computed thus far [**71**]. How do N400 and P600 data bear on compositionality?

3.4. World knowledge. Relevant to this question is an ERP experiment by Hagoort et al. [**79**], using true (4a), false (4b) and semantically anomalous (4c) sentences:[6]

[6]The stimuli were in Dutch and participants were native Dutch speakers.

(4) a. Dutch trains are yellow and very crowded.
b. Dutch trains are white and very crowded.
c. Dutch trains are sour and very crowded.

The words 'white' and 'sour' evoked very similar N400s, in both cases larger than the N400 elicited by 'yellow'. Integrating the meanings of 'white' and 'sour' in the ongoing semantic representation is thus relatively hard. This suggests that, upon encountering 'Dutch trains', features are retrieved which code for the color of Dutch trains – typically yellow-blue – and are responsible for the additional processing costs associated with 'white'.

Although it is notoriously hard to define 'core' semantic features, separating linguistic from world knowledge, it seems nonetheless possible to single out features that are *invariant* across possible contexts of use, as well as across the individuals and communities using the expression in question. That 'sour' cannot be predicated of trains is one such bit of invariant knowledge, and in that sense it is a fact of 'linguistic knowledge'. However, trains differ in color and other properties in space and time, hence that 'white' cannot be applied to Dutch trains reflects a contingent state of affairs which not all users of the expressions 'train' and 'Dutch train' may be aware of ('world knowledge'). The N400 effects found by Hagoort et al. show that 'white' is hard to integrate into its sentence context, which in turn suggests there is some aspect of the meaning of 'Dutch trains' which makes integration hard. This must be knowledge that Dutch trains are yellow and blue, not white. It therefore seems that during processing, meanings are computed – for instance of the compound 'Dutch trains' – that encompass invariant *and* community-specific semantic information, that is, linguistic *and* world knowledge.

As regards compositionality, these data may have two consequences, depending on one's view of the lexicon: either the lexicon encompasses declarative memory in its entirety, and then simple composition seems enough to account for the similarity between the N400 effects, or the lexicon includes invariant meanings only, and then enriched composition – the thesis that the lexicon is not the only source of semantic content – is necessary to explain the observed N400s. If the lexicon is merely a theoretical construct for the linguist's benefit, then it seems one is free to choose one's favorite view, and perhaps salvage compositionality by adopting the first notion. If however 'the lexicon' refers to a natural kind (a functionally segregated component of declarative memory), then enriched composition seems the only option consistent with the empirical data. It is conceivable that brain research will be able to settle the issue of the separability of the lexicon within the human memory system, and thus place constraints on the scope of compositionality with respect to processing data.

3.5. Co-speech gestures. At least one other ERP experiment reporting modulations of the N400 seems relevant to our discussion of compositionality. Özyürek et al. [**155**] showed that N400 effects are elicited by co-speech gestures which do not match the semantic content of the accompanying sentence. This demonstrates that semantic information from different modalities – in this case speech and co-speech gesture – is integrated in the same time frame. The choice mentioned above between two views of the lexicon applies here too. If gesture representations are stored in declarative memory – assuming the latter is entirely contained in the lexicon – then simple composition seems enough to explain the data. However, there exist experimental data showing that subjects attribute different meanings to a given iconic gesture, providing evidence for the non-conventionalized nature of gesture meanings [**20**]. If indeed gesture schemes are not part of the lexicon, some form of enriched composition must occur, which means the semantics of elementary linguistic expressions is not the only source of content for complex meanings. This basic choice between two views of the lexicon, and its consequences for the status of compositionality, illustrates how severe the problem of compositionality's theory-dependence is, and how pressing the need for more realistic constraints on the components of the grammar.

3.6. Fictional discourse. The data we have discussed so far might not constitute direct empirical evidence for or against compositionality, but they reveal the existence of richer meanings that embrace world knowledge or perceptual cues. Experimental research shows that discourse is not only another source of content beyond lexical meanings, but can even modify invariant semantic representations (as defined above) attached to lexical items. One such extreme case of context-sensitivity can be found in fictional discourse. Nieuwland and van Berkum [**148**] show that sentences which are otherwise sensible, like

(5) The peanut was salted.

are perceived as anomalous if they are embedded in a context where the inanimate subject ('the peanut') is attributed animate features. So, in a narrative in which the peanut danced, sang and met an almond it liked, 'salted' resulted in a *larger* N400 compared to 'in love':

(6) The peanut was in love.

This is taken to show that discourse can override even such deep-rooted semantic features as animacy. These findings can also be read as a challenge to Hintikka's principles, for they seem to suggest that meaning is context-dependent, and semantic composition can proceed from the outside-in, that is, from the discourse to lexical meaning.

Processing sentences such as (6) in a fictional context might indeed involve some form of top-down composition which, if it cannot fall back on the mental lexicon as a repository of contextual values of expressions (recall Hodges' argument), then it has to exploit some form of enriched composition, or perhaps give up a share of informational encapsulation – which is the same. But there seems to be another way out for compositionality. One might ask, not what is changed in the meaning of 'peanut' in the context in which it is depicted as animate, but what is preserved of the original, invariant meaning of the word. There is no evidence that any of the original semantic features of 'peanut' are maintained. Therefore, the word form 'peanut' in the fictional context at issue here may just be used as a label for an animate subject or, more precisely, as a proper name with a reference but no (or perhaps very little) sense. This could easily be handled in a compositional semantics. Processing the adjective 'salted', given the plausible combination with 'peanut', might recover its original sense, and this would explain the larger N400. This does not detract from the interest of the data, nor from the explanation proposed by Nieuwland and van Berkum. It does exemplify, however, the kind of problems one encounters when trying to put compositionality to test, and in particular the exceptional elbow room the principle leaves to its own application. It is precisely this resilience which has been taken by many as empirical vacuity.

3.7. Semantic attraction. The strict reading of compositionality implies that only two sources of constraints can interact to produce complex meanings: syntax and the lexicon. A further assumption is that the syntax is an analysis of the sentence *as is given*, for instance a formal decomposition into constituents.[7] Combining lexical meanings (representations or types, depending on one's ontology) based on the syntactic structure will produce an input for interpretation. In this sense, meaning is often said to be dependent on the syntax. Kim and Osterhout [**111**] designed an ERP study to test the extent to which syntax is actually in control of the composition process. They presented participants with sentences like

(7) a. The hearty meal was devouring the kids.
b. The hearty meal was devoured by the kids.

and found that 'devouring' in (7a) elicited a larger P600 compared to 'devoured' in (7b). If the syntax were taken as is given in the sentence – in well-formed sentences, that is – and if it were only proposing input to the semantics, an N400 to 'devouring' should be expected: indeed (7a) is syntactically well-formed, whereas a semantic constraint (animacy) appears to be violated. The P600 indicates that (7a) is perceived as a syntactic violation, originating from the impossibility at 'devouring' of building a passive construction. At the verb's stem

[7] The extent to which a syntactic analysis is allowed to deviate from the surface form of an expression is a matter of considerable debate. For a discussion, see Culicover and Jackendoff [**32, 33**].

'devour-', the passive is the only continuation compatible with a plausible meaning for the sentence, as is testified by (7b). The data therefore show not only that semantic attraction to a more plausible interpretation is an important source of constraints in sentence processing – which could also be concluded if 'devouring' induced a larger N400 – but also that such constraints can override syntactic cues as these are given in the input – which is what the P600, as a syntax-related effect, shows. Compositionality can be salvaged only by assuming that semantic attractors, such as 'kids devour hearty meals', are configurations of the lexical network and not, as would seem more intuitively appealing, the result of inference. But this move is once again paradoxical. Compositionality was introduced to explain productivity, and therefore to *ease* the burden on storage. Now it seems we need a growing inventory of stored semantic notions to maintain compositionality. The same holds, *mutatis mutandis*, for the cases involving world knowledge and co-speech gestures discussed above.

3.8. Coercion. A phenomenon that is often taken as a challenge to compositionality is complement coercion. Consider the following sentences:

(8) a. The journalist began the article after his coffee break.
  b. The journalist wrote the article after his coffee break.

The intuitive difference between (8a) and (8b) is that, while in 'wrote the article' the relevant activity (writing) is mentioned as part of the asserted content, it is not in 'began the article'. So, if a full event sense is to be recovered from (8a), the activity must be inferred based on other semantic cues present in the sentence and stored knowledge. One candidate analysis, due to Pustejovsky [**171**], focuses on the NP as the source of the difference between the two sentence types. On this view, 'the article' is an entity-denoting expression, which combined with verbs such as 'begin' denotes an event. Coercion is seen as an instance of type-shifting, where types specify a basic ontology of entities, events etc. An alternative analysis centers around the VP, and posits event structures of varying complexity depending on the VP's aspectual class [**212**]. Each VP is represented semantically by a quadruple $< f_1, f_2, e, f_3 >$, where f_1 represents a force being exerted, f_2 the object or state driven by the force, e the goal toward which the exertion of the force is directed, and f_3 the state of having achieved that goal. There may be empty slots in the quadruple, again depending on the *Aktionsart* of the VP. Accomplishments such as 'write an article' feature a full event structure, while achievements such as 'begin an article' include only a punctual event e (the beginning of an unspecified activity involving the article) and a consequent state (of having begun that activity). In this analysis, coercion is formalized as the transition to a richer event structure, in which the activity f_1 is also represented.[8] Both proposals rely on some form of enriched composition, as in both cases an operation of meaning assembly that is not syntactic in nature (type-shifting or transition to a richer event structure) is postulated.

An interpretation of (8a) in which the activity remains unspecified is conceivable, and it falls to experimental research to provide evidence for or against the existence of enriched composition in cases of complement coercion. A series of studies have shown that coercing sentences result in increased processing costs compared to controls [**137, 205, 204, 136, 173**], and some ERP evidence for complement coercion will be presented in chapter 6. These data seem to challenge simple composition.[9] A compositional analysis may still be applicable if the operation responsible for generating enriched meanings (for instance, type-shifting) is incorporated into the syntax. However, this choice can be criticized on different grounds. On empirical grounds, because it predicts that 'article' in (8a) would result in a P600 effect, which correlates with syntactic complexity. However, the available data show a different effect than the P600 (see chapter 6 and indirectly the MEG study by Pylkkänen and McElree [**173**]). On theoretical grounds, because a syntactic reduction of coercion requires syntactic representations which resemble less and less a formal decomposition into constituents [**32**]:

[8] This transition is in fact based on inference and unification, as we will show in chapter 6.

[9] Evidence for enriched composition in cases of aspectual coercion has also been found [**163, 162**].

the simplicity and theoretical elegance which are gained by reintroducing compositionality are lost at the level of description of syntactic structure.

4. The balance between storage and computation

In this chapter we have tried to show that compositionality, properly operationalized, can be tested against empirical data. We have also seen that behavioral and neurophysiological data undermine compositionality (simple composition), unless the balance between storage and computation is upset in favor of storage. It now seems appropriate to ask whether there is any interesting sense in which compositionality can be said to hold.

Compositionality (simple composition) remains effective as an explanation of cases in which processing complexity increases due to syntactic factors only. However, it falls short of accounting for situations in which complexity arises from interactions with the sentence or discourse context, perceptual cues, and stored knowledge. The idea of compositionality as a methodological principle is appealing, but imputing the complexity to one component of the grammar or another, instead of changing the notion of composition, is not always an innocuous move that leads to fully equivalent theories. One may be tempted to believe that equivalent theories in this sense are *in principle* indistinguishable in the face of empirical data. Nevertheless, neuroscience grants us restricted selective access to linguistic processes and representations in the brain, as exemplified by the difference between N400 and P600. Therefore, there is at least a chance that what appear to be neutral methodological choices are in fact controvertible given the data. Compositionality sets also an upper bound on the degree of informational encapsulation that can be posited by modular or component-based theories of language: simple composition ties in with a strongly modular take on meaning assembly, which is seen as sealed off from information streams other than the lexicon and the syntax. Empirical data seem to suggest that this upper bound is not always attainable. This implies a weakening of the notion of compositionality, but also more complex traffic either within a module or between modules. So compositionality is also crucial for issues of architecture of, and connectivity within, the language system. Perhaps the most important of these issues is the balance between storage and computation. Compositionality can often be rescued by increasing the demand on the storage component of the architecture, whereas it must be abandoned if one puts more realistic constraints on storage. In the latter case, of course, the demand on the computational component is increased.

Part 2

The electrophysiology of meaning

CHAPTER 4

Processing temporal constraints

This chapter is a modified version of G. Baggio. Processing temporal constraints: An ERP study. In M. Gullberg & P. Indefrey (Eds.), *Time to Speak: Cognitive and Neural Prerequisites for Time in Language*. Wiley-Blackwell, 2008. Appeared also in *Language Learning* 58, 2008: 35-55.

1. Introduction

Reference to time is ubiquitous in natural language, to the point that nearly every assertion involves the location of some event within a temporal coordinate system. Much attention has been paid to the linguistic structures used to encode temporal information, which range from prepositional phrases ('before the dawn') to verb suffixes ('-ed' in the English regular simple past), to expressions borrowed from technical languages, for instance mathematics ('10^{-35} seconds into the expansion phase'). Linguists have provided precise accounts of the functioning of these devices, often using the formal tools made available by modern logic. However, comparatively little is known about how expressions of time are represented and used to compute temporal reference in the brain. Our aim here is to contribute to this line of research by combining a formal analysis of tense with ERP data on tense processing.

1.1. An analysis of tense and temporal adverbials. In this section, we will present a linguistic analysis of sentences in which tense is correctly or incorrectly used, like:

(1) a. Last sunday Vincent *painted* the window frames of his country house.
b. Last sunday Vincent *paints* the window frames of his country house.*

These sentences recognizably form a minimal pair of the kind customarily used in language processing experiments. The formal treatment of tense outlined below is intended to drive the interpretation of the ERP data presented in section 3.2. The main theoretical claim will be that processing sentences such as (1a) involves solving a constraint satisfaction problem where the relevant temporal constraints are introduced by the semantics of the adverbial 'last sunday'[1] and of the verb 'painted'. Conversely, processing a sentence such as (1b) leads to a failure to solve simultaneously such constraints. Even though the analysis developed here brings semantics center stage, tense has ramifications in morphology which can hardly be ignored. Before we turn to semantics, some notes on morphology are therefore in order.

1.1.1. *Morphology.* Comrie [**27**] defined tense as the "grammaticalization of location in time". Unpacked, this definition reads that tense is concerned with the location of events in time – or, possibly, a cognitive representation of time – and that the expression of time is (i) obligatory and (ii) morphologically bound on verbs. Notice, however, that this can only be taken to imply that verbs obligatorily carry temporal information via morphemes, and not that there are morphological rules which determine, for any given verb form, whether tense is correctly used. Let us try to clarify this point with reference to examples (1).

[1]Syntactically, expressions that denote temporal frames, for instance 'last sunday', are NPs and not adverbs. On the other hand, these expressions seem to modify the main verb, and in this sense they function like adverbs. The syntactic analysis presented in 4.1 reconciles these two views by postulating a silent preposition, e.g. 'on', which makes it possible to attach 'last sunday' in the syntactic structure as a prepositional phrase, that is, as an adjunct of time. In what follows, we will call these expressions 'adverbials' to distinguish them from genuine adverbs of time like quantification and frequency adverbs.

The verb forms 'painted' in (1a) and 'paints' in (1b) convey, via the suffixes '-ed' and '-s', the information that the painting event is located, respectively, in the past and in the present of the moment of speech. Importantly, both forms are morphologically correct, in the sense in which, for instance, the over-regularized verb form 'goed' (past of 'to go') is not. What makes 'paints' incorrect in the context of (1b) is a mismatch between its semantics, that is, the fact that it refers (at least by default) to a present event, and that of the adverbial 'last sunday', that defines a time window in the past of the moment of speech in which the main event is to be located.

From these considerations, two views on the processing consequences of the violation in (1b) ensue. On the first, (1b) is perceived by the language system as a semantic violation, and not as a morphological one, because, as we have seen, 'paints' is in fact morphologically well-formed. On the second, 'paints' in (1b) is perceived as a morphological anomaly, even though its origin is semantic. This option requires the additional hypothesis that the system is endowed with an interface component, in Jackendoff's [**99**] sense, that mediates between semantics and morphology, and that, on the basis of the meaning of 'last sunday', constrains the set of suffixes which the verb stem 'paint-' – or any other verb stem, for that matter – can take. In this chapter, we wish to remain neutral with respect to this issue, despite its obvious importance. Instead, we will consider the processing consequences of the observation that, regardless of which hypothesis is correct – that is, regardless of whether (1b) is perceived as containing a semantic or a morphological anomaly – it is the mismatch between the *semantic* features of the adverbial and the verb which determines that tense is used incorrectly.

1.1.2. *Semantics.* In the coming paragraphs, we will present an analysis of sentences like (1a) and (1b), trying to add formal detail to the claim that tense violations are semantic in nature. The analysis is based on a lightweight adaptation of the theory of tense and aspect of van Lambalgen and Hamm [**212**], which has been introduced in previous chapters. The formalism adopted by van Lambalgen and Hamm is an Event Calculus – different though from the event calculi developed in the Davidsonian tradition, for instance by Parsons [**157**] or Larson and Segal [**121**] – reformulated using the computational machinery of Constraint Logic Programming.

The Event Calculus is a many-sorted predicate logic with sorts for eventualities viewed either perfectively, called *events*, or imperfectively, call *fluents*. As we shall see soon, fluents can be used to represent both processes and temporal intervals such as, for instance, the period denoted by 'last sunday'. Temporal and causal relations are modeled using *predicates* and *constraints*, combined in formulae called either *clauses* or *integrity constraints*. Here we will employ two predicates only: $Happens(e, t)$, which means that event type e has a token at time t, and $HoldsAt(f, t)$, meaning that the fluent f holds at time t. Constraints are equalities ($x = y$) or inequalities ($x \gtreqless y$) whose function is to relate the values of the temporal variables that occur as arguments of the predicates.

Discourse comprehension involves the construction of a model making discourse true [**212, 194**]. More precisely, as each word is processed, an *estimate* of the meaning intended by the speaker is computed, that is, a representation making most of what his heard or read true and consistent with the grammar and stored knowledge. This process is incremental, in that each new incoming word results in an update of the model computed during the preceding stage. Lexical meanings can be formalized as sets of *clauses*, of which an example will be given soon (see also the scenario for 'writing a letter' in chapter 2), and updates are regimented by *integrity constraints*. Recall that these are statements that can take the form of obligations '?φ `succeeds`', forcing an update of the model satisfying φ, or prohibitions '?φ `fails`', blocking updates of the model satisfying φ.

The tenses can be treated as integrity constraints which instruct the system to update the current discourse model so as to locate the relevant event in the past, present or future of the moment of speech [**212**]. Further characterization of the event may involve integrity constraints for aspectual profiles (perfect/imperfective; chapter 2). The following integrity constraints may serve as a first approximation of the meaning of the past tense:

(2) a. $?Happens(e, t) \wedge t < now$ `succeeds`
b. $?Happens(e, t) \wedge t = now$ `fails`

While (2a) forces the system to update the current discourse model such that e is located at t, in the past of the moment of speech *now*, (2b) captures the implicature that the event e lies entirely in the past. However, as has often been observed (see the classic Partee [**158**]), the past tense is anaphoric, for it requires an anchor expression, such as a temporal adjunct, in the preceding discourse context. Thus, the sentence 'Vincent painted the window frames of his country house', uttered without a prior temporal context, would be under-informative ('When exactly did he do so?') [**68**]. This suggests that (2) should be refined as follows:

(3) a. $?HoldsAt(f, t) \wedge Happens(e, t) \wedge t < now$ `succeeds`
b. $?HoldsAt(f, t) \wedge Happens(e, t) \wedge t = now$ `fails`

The meaning of the present tense is captured by the following integrity constraint (see van Lambalgen and Hamm [**212**] for a more rigorous analysis):

(4) $?HoldsAt(f, t) \wedge Happens(e, t) \wedge t = now$ `succeeds`

The fluent f in (3) and (4) has to be unified with material from the preceding context, for example a temporal adverbial. 'Last sunday' can be analyzed as proposed by Hamm et al. [**83**]. We first partition the time line in seven segments of equal length, corresponding to the days of the week: $f_{Su}, f_{Mo}, ..., f_{Sa}$. Next we stipulate that the meaning of the adverbial is given by the clause (5), where f_{CPSu} denotes the closest past sunday:

(5) $HoldsAt(f_{Su}, s) \wedge \; s < now \; \wedge |now - s| \leq 7 \; days \rightarrow HoldsAt(f_{CPSu}, s)$

Unification works as follows. Upon encountering a sentence like (1a) or (1b), the system is incrementally confronted with a constraint satisfaction problem. Let us first consider (1a). Processing the phrase 'last sunday' results in a model in which the query $?HoldsAt(f_{CPSu}, t)$ succeeds, that is, in which the antecedent of (5) holds. When the verb 'painted' is processed, the system tries to satisfy (3). This is done by unifying f in (3) with f_{CPSu} in (5) (f=f_{CPSu}) and t in (3) with s in (5) ($s = t$). All constraints set up by (1a) can be satisfied, as no inconsistency follows. Let us now turn to (1b). Processing the phrase 'last sunday' leads again to a model in which the antecedent of (5) holds. When the verb 'paints' is processed, the system tries to satisfy (4). However, unifying s with t ($s = t$) would lead to an inconsistency. While $s < now$ has been satisfied when processing the adverbial, (4) requires on the contrary that $t = now$ (or equivalently, under the proposed unification, $s = now$) succeeds. These two constraints cannot be satisfied simultaneously. This leads to the following definition of 'tense violation': there is a mismatch between an temporal modifier and verb's tense, or two verbs' tenses, when the unification of the temporal variables set up by the relevant expressions fails.

1.1.3. *Consequences for ERP studies of tense.* The definition just proposed can be used in a critical assessment of the ERP literature on tense. The first attempt at investigating tense violations using ERPs was made by Kutas and Hillyard [**116**]. In this experiment, a group of English speakers was presented with sentences like the following:

(6) a. Most of the earth's weather happens in the bottom layer of the atmosphere *called* the troposphere.
b. Most of the earth's weather happens in the bottom layer of the atmosphere *calls* the troposphere.*
c. The eggs and meat of this turtle are *considered* choice food by many people.
d. The eggs and meat of this turtle are *consider* choice food by many people.*
e. This allows them to *stay* under water for a longer period.
f. This allows them to *stayed* under water for a longer period.*

Incorrect sentences elicited a positive shift peaking at about 300 ms following the verb and approaching significance at parietal sites, and a negative wave in the 300-400 ms window, suggestive of an N400 effect [**117**]. Leaving a functional account of these effects aside, there are two issues with this study. First, (6b), (6d) and (6f) are not instances of tense violations, as none of these sentences contains a temporal expression, such as a temporal adverbial or a temporal preposition phrase, with which the verb's tense fails to agree. Second, anomalies are realized in rather different ways: in (6b) the present indicative 'calls' replaces the past participle 'called'; in (6d) the infinitive 'consider' replaces the past participle 'considered'; and in (6e) the past participle 'stayed' replaces the infinitive 'stay'.

Osterhout and Nicol [**154**] investigated the ERP correlates of "verb tense violations" in modal constructions, presenting to a group of English speakers sentences like:

(7) a. The cats won't *eat* the food that Mary leaves them.
b. The cats won't *eating* the food that Mary leaves them.*
c. The expensive ointment will *cure* all known forms of skin disease.
d. The expensive ointment will *curing* all known forms of skin disease.*
e. The new fighter planes can *fly* faster than anyone had expected.
f. The new fighter planes can *flying* faster than anyone had expected.*

In the 300-500 ms window, incorrect sentences were more positive at posterior sites of the midline, whereas the LAN (Left Anterior Negativity) elicited by anomalous verbs did not differ significantly from the effect evoked by correct verbs. In the 500-800 ms time interval, incorrect sentences were more positive at mid-line and posterior sites, consistently with the distribution of the P600 [**151, 78**]. The ERPs elicited by sentence-final words in the violation condition were more negative-going compared to those observed in correct sentences, they started 200 ms following the sentence-final word and continued throughout the epoch. Also in this case, it should be noted that sentences (7) do not contain tense violations. The verbs lack an anchor point in the preceding discourse with respect to which tense agreement can be evaluated.

In an ERP study by Allen et al. [**1**], "syntactic (tense) violations" were investigated in sentences containing either high-frequency (HF) or low-frequency (LF) verbs:

(8) a. The man will *work* on the platform. (HF)
b. The man will *worked* on the platform.* (HF)
c. The man will *sway* on the platform. (LF)
d. The man will *swayed* on the platform.* (LF)

Correct low-frequency verbs elicited a larger N400, ungrammatical verbs a larger P600, and low-frequency anomalous verbs a bi-phasic increase of N400 and P600. In the 500-900 ms window a significant main effect for grammaticality was found, maximal on posterior sites as is typical of the P600. Unfortunately, (8*) do not contain tense violations, contrary to the authors' claim that the "sentences 'He will walked' and 'He will swayed' are equally and unconditionally ill-formed with respect to tense" [**1**].

In brief, the experiments of Kutas and Hillyard [**116**], Osterhout and Nicol [**154**] and Allen et al. [**1**] did not bring into play genuine tense violations. In none of these studies stimulus sentences were such that the anomalous verb located the event described by the main clause outside a frame specified by an anchoring temporal expression.

Genuine tense violations have been investigated in two ERP studies. Fonteneau et al. [**56**] presented a group of French speakers with sentences like:

(9) a. Demain l'étudiant *lira* le livre.
b. Demain l'étudiant *lisait* le livre.*

The adverbial modifier 'demain' specifies a temporal frame within which the eventuality denoted by the main clause is taken to occur. Because the lapse of time denoted by 'demain' is located after the moment of speech, the tense of the verb should be future as in (9a). In (9b) the verb 'lisait' locates the event in the past, thus outside the temporal frame specified by 'demain'. Based on the definition given above, (9b) is therefore a genuine tense violation. Anomalous verbs evoked a bi-phasic wave in the 450-550 ms interval following the onset of the critical word, with a negative maximum over posterior sites and a positive anterior peak. Given current knowledge of ERP components as modulated by linguistic processes, it is hard to make sense of these effects, for they do not fall within any of the (E)LAN, N400 or P600 classes. Also, given the semantics of the Imparfait and its cognitive-computational consequences [**212**], the Passé Simple 'lut' might have been a more suitable candidate for realizing tense violations without aspectual confounds.

Steinhauer and Ullman [**192**] presented their subjects with sentences like:

(10) a. Yesterday, I *sailed* Diane's boat to Boston.
b. Yesterday, I *sail* Diane's boat to Boston.*

In (10), 'yesterday' specifies a past time frame for the occurrence of the main event. In (10b), 'sail' locates the event in the present, thus outside the period denoted by 'yesterday'. In the 400-900 ms time window following verb onset, tense violations elicited a consecutive LAN (400-500 ms) and P600 (600-900 ms), taken by the authors as signatures of morpho-syntactic processing. Regardless of whether 'sail' was perceived as semantically or morphologically anomalous (see 1.1.1), the LAN observed by Steinhauer and Ullman indicates that *semantic* information, in the form of temporal constraints – or anything functionally equivalent – set up by the adverbial, is used by the system as early as 400 ms after verb onset. The results of the experiment reported below suggest that the ERP effects of tense violations might be even earlier than this.

2. Method

2.1. Materials. The Dutch materials used in the experiment were 80 critical sentences, 40 correct and 40 tense violations, and 240 fillers. All sentences had the following structure: a past temporal adverbial was followed by the main, simple past or present tensed verb, a subject NP and an object NP, in most cases modified by a PP. Below are some examples of the stimuli used:

(A) Afgelopen zondag *lakte* Vincent de kozijnen van zijn landhuis.
Last sunday *painted* Vincent the window-frames of his country-house.
'Last sunday Vincent *painted* the window frames of his country house.'
(B) Afgelopen zondag *lakt* Vincent de kozijnen van zijn landhuis.*
Last sunday *paints* Vincent the window-frames of his country-house.
'Last sunday Vincent *paints* the window frames of his country house.'

The temporal adverbials were 'Vorige week' ('last week'), 'Vorige maand' ('last month'), 'Vorig jaar' ('last year'), 'Vorige eeuw' ('last century'), 'Afgelopen N' for each day of the week, each month and each season. Each adverbial was used at most three times. Eighty verbs were used, 40 regulars and 40 irregulars, 20 activities, 25 accomplishments and 35 achievements. The mean length (correct: M=7.41, SD=2.56; violation: M=6.79, SD=2.25) and raw word frequency (correct: M=6898.34, SD=34220.5; violation: M=7174.91, SD=44939.35) of the verbs were normed using the CELEX corpus for Dutch [**4**]. There were no differences between conditions (T-tests, $P > 0.9$ in all comparisons). All non-critical words, including sentence-final words, and sentence length were the same across conditions as illustrated by (A)-(B). Fillers were 160 grammatically well-formed sentences containing the prepositions 'after' and 'before' and 80 verb-adverbial sentences (40 correct and 40 tense violations) similar to critical items but constructed using different lexical material. Two test versions were constructed, consisting of pseudo-randomized lists of critical and filler items.

2.2. Participants. Twenty five students participated in the experiment. One was left out of the final analysis due to a high number (> 20%) of trials contaminated by artifacts, so there remained 24 participants (mean age 24.6, 14 female; with no history of neurological, psychiatric or cognitive disorders). Subjects were selected from the database of the Donders Centre for Cognitive Neuroimaging at the Radboud University Nijmegen. They received € 6 per hour or course credits for taking part in the experiment.

2.3. Procedure. After applying the electrodes (see 2.4), participants were conducted into the experimental room and were asked to sit in front of a video monitor. The stimuli were presented on the screen word-by-word (600 ms SOA, 300 ms word duration, white on black background), and were followed by a fixation cross presented for 1500 ms. Participants were instructed to read each sentence carefully and to blink or move only when the fixation cross was shown. The experiment took about 50 minutes to be completed and was divided into 4 blocks of 80 trials each.

2.4. Recording. The EEG/EOG was recorded from 32 sintered Ag/AgCl electrodes. Two electrodes were placed at the outer canthi of the left and right eyes. One electrode below the left eye monitored vertical eye movements. The remaining 29 electrodes were arranged according to American Electrophysiological Society conventions at the following locations: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, TP10, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2. The left mastoid electrode served as true reference. All electrodes were re-referenced off-line to a linked mastoid. Electrode impedance was kept below 5 kΩ throughout the experiment. The EEG/EOG was amplified by a multichannel BrainAmp DC system, with the following settings: 500 Hz sampling rate, a low pass filter at 70 Hz, and a 10 s time constant.

2.5. Data Analysis. The data was analized using FieldTrip,[2] a MATLAB package for processing EEG/MEG data. Several transforms were applied to each participant's data-set. Segments corresponding to the verb and the sentence-final word were extracted from the EEG with an interval of 200 ms before and 1000 ms after stimulus onset. Baseline correction was based on a 200 ms pre-stimulus interval. Two FieldTrip functions were used for artifact rejection. The first rejected all trials containing activity exceeding a threshold of $\pm 100\ \mu V$. The second function discarded trials contaminated with eye movements or blinks based on thresholding the z-transformed value of the raw data in the EOG channels, preprocessed using a band-pass filter at 1-15 Hz. A 30 Hz low-pass filter was applied to the segmented, artifact-free data. ERPs were obtained for each participant by averaging over trials in each experimental condition. A 5 Hz low-pass digital filter was used to produce the waveforms shown in figures 1-2. Topographical plots and statistical analyses used the 30 Hz low-pass filtered data. Statistical analyses were based on a non-parametric randomization procedure **[129, 130]** which took as input mean amplitude values in each condition in time bins of 100 ms, starting from the onset of the relevant word and ending 1000 ms after, and returned a cluster of electrodes in which the difference between the conditions was significant in each time bin, the sum of T statistics in that cluster, and Monte Carlo estimates of P-values.

3. Results

A visual inspection of ERP waveforms elicited by the verb (figure 1-b) reveals a negative deflection that peaks at approximately 100 ms after verb onset, and is followed by a positive shift with a trough around 200 ms. The amplitude of these two components, referred to as N100 and P200 respectively, in not affected by the experimental manipulation: there are no electrode clusters between 0 and 200 ms after verb onset at which the mean amplitude difference between tense violations and correct sentences is significant (table 1). The P200 is followed by a negative shift, larger for tense violations over left-anterior scalp sites, starting around 200 ms from verb onset and lasting for approximately 200 ms (figure 1-b). There are clusters of electrodes in which the negativity is significant between 200 and 400 ms (table 1, figure 1-a), and marginally significant between 400 and 500 ms. We refer to this effect as a left-anterior negativity (LAN). Around 600 ms following verb onset, the waveforms are characterized by a bi-phasic (positive-negative) response (figure 1-b), which however is not different between conditions in either the 500-600 ms or the 600-700 ms intervals (table 1). A positive shift, larger for tense violations at right-posterior electrodes sets on at about 700 ms and lasts for the entire epoch (figure 1). However, only marginally significant positive clusters were found between 700 and 1000 ms (table 1). This effect is an instance of P600 **[151, 78]**.

[2]For more information, see `http://fieldtrip.fcdonders.nl/`

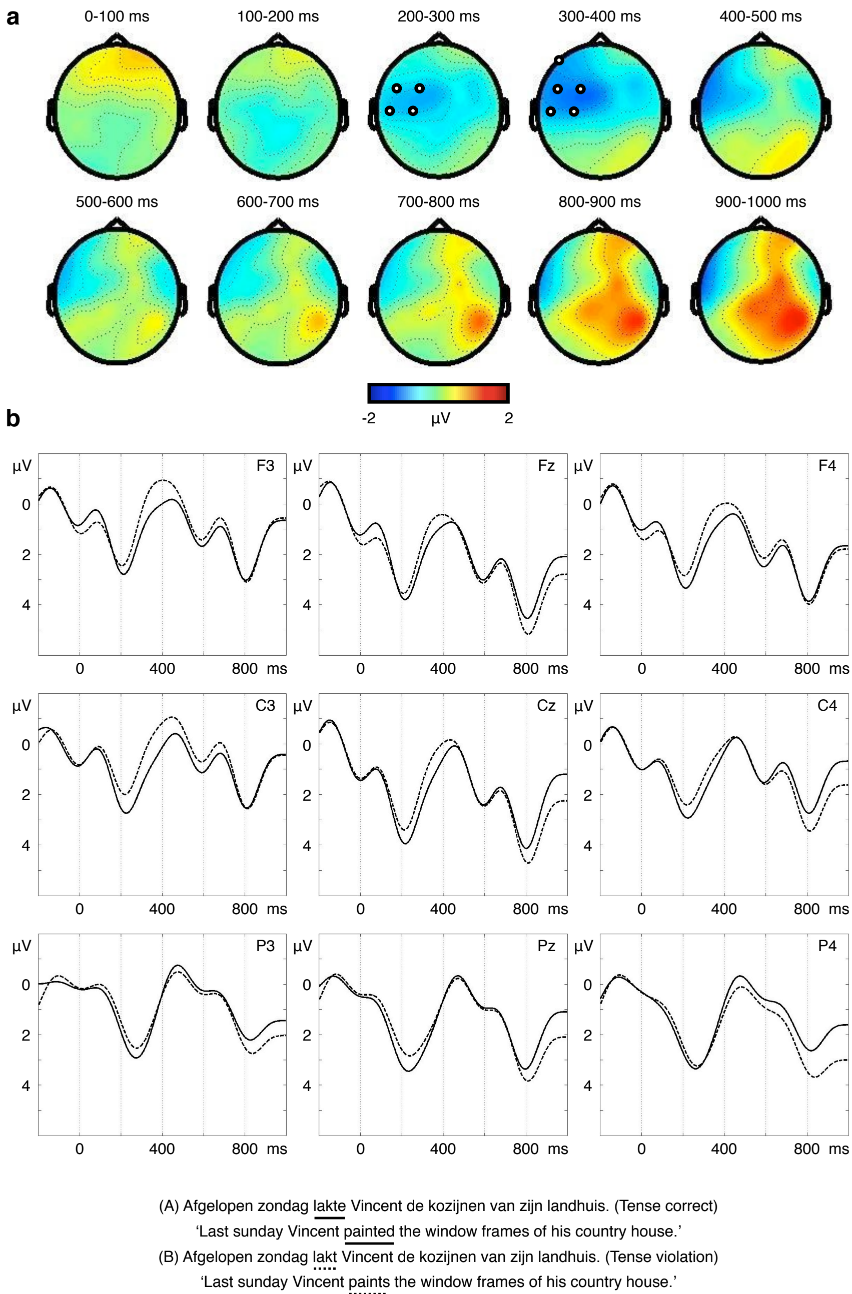


FIGURE 1. (a) Grand-average (N=24) topographies displaying the mean amplitude difference between the ERPs evoked by the verb in tense violations and in correct sentences. (b) Grand-average (N=24) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the verb in tense violations and correct sentences. Negative values are plotted upwards.

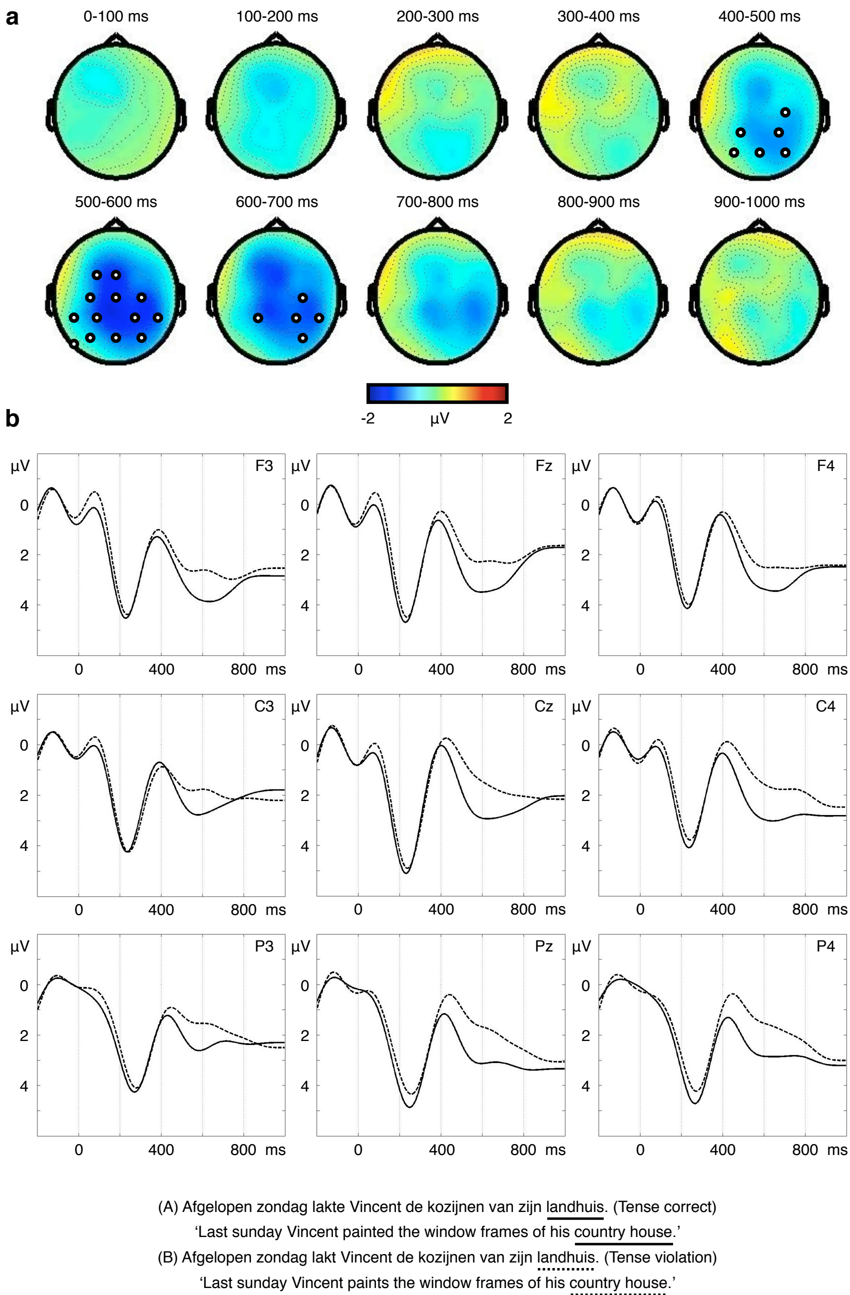


FIGURE 2. (a) Grand-average (N=24) topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final word in tense violations and correct sentences. (b) Grand-average (N=24) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the sentence-final word in tense violations and correct sentences. Negative values are plotted upwards.

The waveforms evoked by sentence-final words show similar N1-P2 complexes (figure 2-b). The amplitude of these early, largely endogenous components is not affected by verb tense (table 1). The P2 is followed by a negative shift, peaking at about 400 ms following the onset of the sentence-final word (figure 2-b). There are no significant clusters between 200 and 400 ms, but the negativity becomes significantly larger for tense violations between 400 and 700 ms (figure 2, table 1). The effect bears some superficial resemblance with the N400, in particular in its polarity and central distribution [**117**]. However, it has a later maximum and is more sustained compared to the N400. In what follows, we will refer to this effect as sentence-final negativity (SFN).

Time	Verb		Sentence-final word	
	Negative clusters	Positive Clusters	Negative clusters	Positive Clusters
200-300 ms	$T(22) = -9.97$ $P = 0.046$			
300-400 ms	$T(22) = -14.46$ $P = 0.026$			
400-500 ms	$T(22) = -7.81$ $P = 0.056$		$T(22) = -15.33$ $P = 0.022$	
500-600 ms			$T(22) = -34.62$ $P = 0.006$	
600-700 ms			$T(22) = -6.66$ $P = 0.038$	
700-800 ms		$T(22) = 5.01$ $P = 0.09$		
800-900 ms		$T(22) = 4.44$ $P = 0.064$		
900-1000 ms		$T(22) = 6.70$ $P = 0.05$		

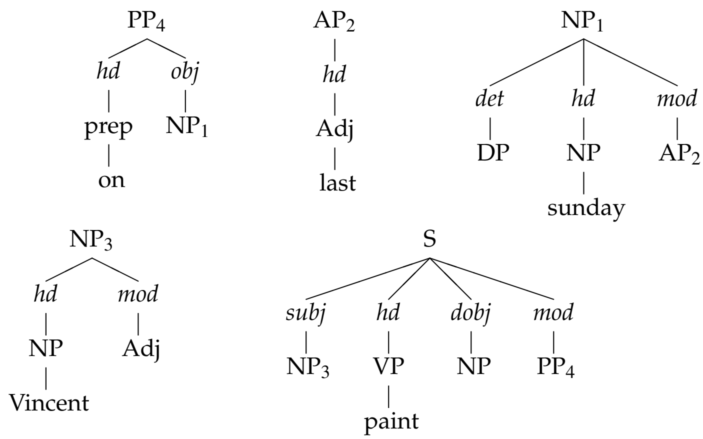
TABLE 1. Summary of cluster-based T-statistics for the ERP data. Tense violations and correct sentences are compared at the verb and at the sentence-final word in time bins of 100 ms starting from word onset. The first significant effects occurred at 200-300 ms. Empty cells denote the absence of (marginally) significant clusters.

4. Discussion

As the reader may recall from chapter 1, relations between linguistic theory and processing data are notoriously problematic [**165**]. In particular, the notion that behavioral and brain data can constrain linguistic theories, and that the theory of 'competence' can be shaped by 'performance' data, seems far from accepted, even though it is sometimes granted that the converse is the case [**99**]. Following Jackendoff [**101**], we assume that a linguistic theory can be chosen over its competitors also based on its capacity to account, at the appropriate level of analysis [**131**], for data on language processing and acquisition. Here we will outline a theory of tense at the syntax-semantics interface that meets this requirement. The model is moreover consistent with the observation that the constraints set up by a temporal adverbial are used to process a main verb as early as 200-300 ms after the onset of the latter.

4.1. Tense, parsing and comprehension. A missing piece in the analysis of temporal expressions in section 1.1.2 is syntax. Here we adopt a lexicalist parser developed by Vosse and Kempen [**221**], and applied by Hagoort [**71**] to account for syntax-related ERP effects. The parser assumes that syntactic information is stored in the mental lexicon as three-tiered unordered trees ('syntactic frames') that specify the admissible syntactic environment of a particular word. The syntactic frames associated with 'last', 'sunday', 'Vincent' and 'paint' are shown in (11), and 'on' is assumed to be silent as in (1). The topmost tier of each frame is constituted by a single *root node* which contains phrasal information (NP, S etc.) and which dominates a set of *functional nodes* in the second tier. For instance, an S node dominates four functional nodes, that is, *subj*ect, *head*, *d*irect *obj*ect and *mod*ifier; an NP node will have at most three functional nodes, namely *det*erminer (if it is not associated with a proper name), *h*ead and *mod*ifier; and so forth. The third tier contains again phrasal nodes, called *foot nodes*, to which the root node of another syntactic frame can be attached.

(11)



Parsing consists in (i) checking agreement features of words such as number, gender, tense etc., (ii) checking word order, and (iii), only if these checks are passed, in unifying frames with identical root and foot nodes. Subscripts in the labels of root and foot nodes are used to indicate unification links in (11). Unifying root and foot nodes with the same subscripts yields the syntactic structure of (1) up to and including the verb 'painted/paints'.

Vosse and Kempen [**221**] assume that functional nodes carry features upon which the agreement checking (i) is based. For example, the *subj*ect node of 'paints' carries the features *Case=nominative*, *Person=third* and *Number=singular*. These are compatible with the features carried by the root node of the frame associated with 'Vincent'. Therefore, the check yields a positive result and the two frames can be unified. Tense could be treated analogously using an additional level of features in the frames of verbs and temporal prepositions. However, the resulting feature structure would then have to be interfaced with the semantics of verbs and temporal adverbials to ensure that the value of the *Tense* feature is consistent with that of the temporal variables in (3)-(5). This implies that three levels of representation have to be coordinated: the syntactic frame, the feature structure, and the semantic structure.

A more parsimonious option is to handle the check for temporal agreement within the semantics. Consider (1a). Suppose that the parser has constructed a syntactic representation for the adverbial 'last sunday', which amounts to unifying the root and foot nodes NP_1 of 'on' (which we assumed is silent) and 'sunday', and the root and foot nodes AP_2 of 'sunday' and 'last'. Moreover, suppose the system has rapidly constructed a semantic representation for the adverbial 'last sunday', which amounts to forcing the query $?HoldsAt(f_{CPSu}, t)$ to succeed. This leads to a minimal model in which the antecedent of (5) holds (recall section 1.1.2). Note that this model satisfies the constraint $s < now$. When 'painted' is encountered, the syntactic frame for 'paint' is retrieved from declarative memory. The root and foot nodes

PP_4 of 'last sunday' and 'paint' are unified as soon as the agreement check is passed. This is carried out within the semantics, by simply updating the initial model according to the integrity constraints (3). Because the unification $s = t$ can be made to succeed, the temporal agreement check is passed, and the two frames can be unified. Consider now (1b). In this case, when 'paints' is encountered, the initial model is updated according to (4). Whereas the adverbial satisfies the constraint $s < now$, the verb satisfies $t = now$, which entails that s and t cannot be unified. The agreement check (i) is not passed, which results in a failure to unify the syntactic frames of 'last sunday' and 'paints'.

4.2. A functional account of the LAN. Negative shifts in the ERPs different from the N400 [**117**] have been reported by several studies. These effects are typically referred to as (early) left anterior negativities, or (E)LANs. The onset of (E)LANs varies as a function of stimulus characteristics, ranging between 150 and 300 ms [**61**]. ELANs (150-200 ms) have been observed in response to rapidly detectable violations of word-category. LANs (300-500 ms) are elicited by morphological anomalies involving case, number and gender [**61, 71**].

The theory of tense at the syntax-semantics interface presented above seems capable of accounting for the early response to tense violations. One key observation in this regard is that an effect which sets on between 200 and 300 ms after the onset of the stimulus arguably does not reflect feed-back or recurrent computations. Rather, it is most probably related to the disruption of a largely automatic, feed-forward spread of activation from sensory areas (visual cortex), through brain regions subserving semantic memory (left temporal cortex), towards more anterior language areas (left frontal cortex) [**72**]. Hagoort [**71**] suggested that (E)LANs reflect the failure to bind two syntactic frames as a result of a negative outcome of the agreement check, or of a failure to find two matching root and foot nodes. The LAN observed in the ERP experiment reported here can be taken as reflecting the failure in the tense agreement check, that is, a failure to solve simultaneously the temporal constraints set up by the adverbial and the verb. This account appears consistent with the time-course of the LAN, for it only requires that the constraint $t = now$, associated with the tense suffix of the verb 'paints', is accessed in semantic memory. From this, given the constraints set up by the adverbial in the preceding processing stages, it follows that $s = t$ (a condition of temporal coherence within the sentence) cannot be satisfied. Note that this account does not require (i) that the meaning of the verb 'paints' is reconstructed from semantic memory, as only the semantic contribution of the tense suffix '-s' is necessary, and (ii) that any form of syntactic structure assembly takes place, because the constraint satisfaction stage actually precedes the unification of syntactic frames.

Although in-depth experimentation and modeling are lacking, the temporal profile of the LAN suggests that checking the satisfiability of a set of temporal constraints – or at least detecting an inconsistency – may have a purely feed-forward neural implementation. This can be contrasted with the (re)computation of a discourse model, which requires a recurrent network architecture [**91, 194**] and gives rise to qualitatively different neurophysiological effects, as we shall see in the next chapter.

4.3. A functional account of the SFN. A number of studies have reported negative shifts in the ERPs in response to sentence-final words in syntactically ill-formed sentences, even when the anomalous word does not occupy the sentence-final position [**151, 78, 152, 153, 139, 149, 150**]. Osterhout and Nicol [**154**] reported negativities in response to sentence-final words preceded by semantic or syntactic violations. Moreover, at some electrode sites, the SFN elicited by doubly (syntactically and semantically) anomalous sentences was an additive function of the SFNs evoked by syntactic and semantic violations.

Osterhout and Holcomb [**151**] argued that the SFN may be an electrophysiological marker of either (i) the perceived ungrammaticality of the sentence, or (ii) the system's effort to find an acceptable syntactic structure for the sentence, or (iii) the semantic, message-level consequences of the sentence-internal violation. As for the SFN elicited by tense violations, (i) seems to imply either that (a) the incorrect use of tense was not perceived when the verb was encountered, but only at the sentence-final word, or that (b) the ill-formedness of the sentence (as opposed to the VP) was perceived at the sentence-final word, as only then it becomes evident that the entire sentence, which consists of a single VP, is ungrammatical. Both these implications seem to be untenable. The LAN elicited by the verb indicates that the violation was in some sense perceived, which rules out (a). Furthermore, any sentence containing an ungrammatical phrase is itself ungrammatical, which makes (b) untestable. The second hypothesis (ii) implies that, at the sentence-final word, the system computes a plausible syntactic representation for the anomalous sentence (1b), eventually unifying the root and foot nodes PP_4 of 'last sunday' and 'paint'. However, computing a new syntactic representation does not solve the sentence-internal problem, which is semantic in nature. Either at the sentence-final stage syntactic unification does away with the agreement check, and then unifying frames does not guarantee that tense and other agreement features are consistent across the representation, or the agreement check is still required, and then it is the semantic constraints that must be readjusted.

This brings us to (iii), according to which the SFN reflects either (a) the disruption of a process involved in computing a model of the sentence, or (b) the attempt of the system to construct a model in which all constraints can be satisfied. One key feature of the Event Calculus is that satisfying a set of constraints actually produces a model which verifies the linguistic material that is given as input [**212**]. Thus, in this framework, the LAN reflects the failure to unify s and t (adding the constraint $s = t$), and therefore indexes the disruption of the processes which would have otherwise led to a model. This rules out (a). According to (b), the SFN would reflect the readjustment of the constraints set up by the verb in order to make the full set of constraints satisfiable. In semantic terms, this can be characterized as forcing the integrity constraints (3) also in the present tense case (1b), yielding a 'narrative present' reading of the sentence. In neural terms, reworking a set of constraints to produce the desired output can be seen as an instance of perceptron learning via back-propagation [**179**], an operation which requires a recurrent architecture. Such feed-back process seems consistent with the time-course of the SFN (400-700 ms), especially in light of the hypothesis proposed above that the LAN (200-400 ms) reflects a feed-forward computation.

CHAPTER 5

Computing and recomputing discourse models

This chapter is a modified version of G. Baggio, M. van Lambalgen & P. Hagoort. Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language* 59, 2008: 36-53.

1. Introduction

In the past three decades, experimental research using event-related potentials (ERPs) has provided numerous insights into word, sentence and discourse comprehension. However, as has been noted, "a cognitive neuroscience approach to language has not as yet merged with linguistic and psycholinguistic research programmes" [**13**]. One research program in linguistics that may contribute to understanding the basis of meaning in the human brain is semantic theory. Logicians and formal semanticists since the 'dynamic turn' [**160**] have shifted their attention from describing semantic competence to modeling cognitive update and information exchange. An example is a recent proposal by van Lambalgen and Hamm [**212**] which regards comprehension as an incremental, yet non-monotonic process whereby temporary structures are set up in working memory, and may be later revised on the basis of additional discourse information. Although there is some evidence for semantic reanalysis [**17, 197**], the issue has received less attention than it deserves. The aim of this chapter is to contribute to filling this gap. We used ERPs to test the recomputation hypothesis presented in chapter 2.

1.1. ERP research on semantic processing. Event-related brain potentials have proved useful to address a number of issues concerning the relative complexity and time-course of semantic processes. Kutas and Hillyard [**117**] conducted the first ERP experiment in which linguistic factors were successfully manipulated, in this case the semantic plausibility of a word given the preceding sentence context:

(1) a. The officer shot the man with a gun.
b. The officer shot the man with a moon.

Compared to 'gun', the anomalous noun 'moon' resulted in a larger negative shift starting around 250 ms after word onset, peaking at 400 ms, and lasting for approximately another 150 ms. This ERP component, called N400 because of its polarity and peak latency, is known not to be affected by other unexpected events, such as variations in the physical properties of the stimuli. Larger N400s are also triggered by semantically plausible words which are judged as less preferred in a given sentence context [**118, 76**], for example 'pocket' in (2b):

(2) a. Jenny put the sweet in her mouth after the lesson.
b. Jenny put the sweet in her pocket after the lesson.

The amplitude of the N400 is also modulated by lexical items which provide information conflicting with the discourse context [**208, 209**] or world knowledge [**79**]. In sum, although every content word evokes an N400, the amplitude of the negative shift is affected by the degree of semantic fit of a lexical item with the preceding context and the knowledge base relevant for its integration.

Semantics-related negative shifts different from the N400 have also been reported. Van Berkum et al. recorded ERPs while participants read [**207**] and listened to [**206**] discourses in which a particular NP in a target sentence could denote either a single referent introduced in the preceding discourse, or two equally suitable referents. For instance (3c), containing the NP 'the girl', could follow either the single-referent context (3a) or the double-referent context (3b):

(3) a. David had told the boy and the girl to clean up their room before lunch time. But the boy had stayed in bed all morning, and the girl had been on the phone all the time.
b. David had told the two girls to clean up their room before lunch time. But one of the girls had stayed in bed all morning, and the other had been on the phone all the time.
c. David told the girl that ...

Referentially ambiguous NPs, such as 'the girl' in (3c) following (3b), elicited a sustained anterior negativity (SAN) that emerged 300-400 ms after noun onset and lasted for several hundreds of milliseconds. The SAN differed from typical instances of the N400 in duration ('sustained') and scalp distribution ('anterior'). Following earlier proposals [**142, 146**], the time-course and the topographical distribution of the observed ERPs are taken to suggest that "at least some of the processing consequences of referential ambiguity may involve an increased demand on memory resources" [**206**].

An ERP study by Münte et al. [**147**], which we discussed in light of formal semantics in chapter 1, also reported sustained anterior negativities. ERPs were recorded while subjects read narratives differing in the initial temporal connective:

(4) a. After the scientist submitted the paper, the journal changed its policy.
b. Before the scientist submitted the paper, the journal changed its policy.

'Before' sentences elicited a larger sustained negativity, maximal over left anterior sites. At the left frontal electrode, ERP responses to 'before' and 'after' diverged approximately 300 ms after sentence onset. The effect lasted throughout the sentence, and was largest during the second clause. The anterior negativity difference between 'before' and 'after' sentences was positively correlated with participants' working memory span. Münte et al. argue that the slow negative shift evoked by 'before' sentences reflects working memory operations involved in computing a model for (4b) in which the events are represented in their actual order of occurrence. That is, in contrast with (4a), (4b) requires *additional* memory resources as the two events are mentioned in reverse temporal order.

The connection between sustained anterior negativities and working memory is made explicit in the papers of van Berkum et al. [**207, 206**] and Münte et al. [**147**]. However, there is as yet no agreement upon a functional account, based on linguistically-informed notions, of these findings. For instance, while van Berkum et al. suggest that the sustained anterior negativity reflects 'referential processing', Münte et al. would rather argue that 'additional discourse-level computations' of the temporal and causal profiles of the events described by 'before' and 'after' sentences are responsible for the observed slow potentials. Matters are further complicated by the finding that sustained anterior negativities are also elicited by constructions in which complexity at the syntax-semantics interface is increased, as in long-distance *wh*-dependencies [**112, 146, 53, 48, 161**].

Despite the differences between the conditions in which sustained anterior negativities have been observed, candidate functional processes can be brought under a single umbrella term, which we shall refer to as 'computing a discourse model'. Formal semantics, at least since Discourse Representation Theory (DRT) [**107**], has assumed that interpreting definite and indefinite NPs, resolving anaphoric pronouns, determining event order, establishing long-distance *wh*-dependencies, and other cross-clause and cross-sentence processes concur in the construction of a discourse model, that is, a cognitive representation making a given narrative true. More recent proposals, which build upon DRT and add some sophistication to it, regard discourse comprehension as a process in which lexical meanings, context and world knowledge interact to produce discourse models [**212, 83**]. Pragmatic constraints and causal/world knowledge are brought to the fore by these accounts. Furthermore, discourse models as envisaged by the theory (called 'minimal models') can be efficiently computed by artificial neural networks, which account for some capabilities and limitations of working memory [**194**]. Therefore, in this framework it becomes possible to raise and to some extent address a number of issues concerning the complexity of computing discourse models in working memory. To see in some detail how this could be done, we must first introduce the linguistic phenomenon with which we shall be concerned, giving a précis of the treatment of the imperfective paradox presented in chapter 2.

1.2. The imperfective paradox. Verb phrases (VPs) can be semantically classified as states ('know', 'love' etc.), activities ('run', 'write' etc.), accomplishments ('write a letter', 'bake a cake' etc.), achievements ('finish', 'reach' etc.), and points ('flash', 'hop' etc.) [**190**]. Accomplishments always involve some activity, from which they can be derived by adding a direct object. For instance, the accomplishment 'write a letter' is constituted by the activity 'write' and the direct object 'a letter', which need not refer to an existing entity, but provides information about the goal toward which the writing activity is directed. We use the term 'activity' to denote both the aspectual class of VPs such as 'write' in the above classification, and the atelic (i.e. non-goal-directed) process involved in accomplishments. 'Progressive' and 'imperfective' will be used interchangeably here to allow readers to see the connection between the semantics of the progressive and the imperfective paradox, although this is not entirely correct [**26**].

Let us consider accomplishments first:

(5) The girl was writing a letter when her friend spilled coffee on the tablecloth.

From (5) the reader would typically conclude that, barring unforeseen circumstances, the girl will attain the desired goal and would thus assent to the statement 'The girl has written a letter' (see chapter 2 and below for evidence supporting this claim). Such an inference is based on the assumption that spilling coffee on the tablecloth is usually *neutral* with respect to the writing activity, that is, it is not a typical proximate cause leading to the termination of the activity. It is quite possible to imagine situations in which writing was temporarily interrupted or even terminated by the accident. However, as is shown by the data reported in sections 2.2.2 and 3.1, failing to explicitly mention an obstacle in the discourse is sufficient to lead the reader to assume that there was no such obstacle to attaining the intended goal.

We hypothesize that the inference to a goal state is *defeasible* or *non-monotonic*, that is, it can be suppressed if the discourse describes an event which terminates the relevant activity:

(6) The girl was writing a letter when her friend spilled coffee on the paper.

Assuming that writing was intended to occur on the same paper sheets upon which coffee was spilled, the accident is sufficient to terminate the activity and it is therefore a *disabling* condition for obtaining a complete letter. Accordingly, on the basis of (6) the reader would assent to 'The girl has written no letter'.

Suppression can obtain only with accomplishments, not with activities [**180**]. In most accomplishments, the object NP 'a letter' expresses the existence of a natural culmination point, or 'canonical goal', toward which the writing activity is directed, namely a complete letter. Activities, for instance 'writing letters', do not involve any such canonical goal. Here the use of the bare plural 'letters' indicates that the number of letters is (for the speaker and the hearer) unspecified and that, as a result, the activity has no natural culmination point. A narrative containing the activity VP 'writing letters' will therefore be interpreted as entailing that 'The girl has written one or more letters', regardless of the consequences of the second event on the writing activity:

(7) The girl was writing letters when her friend spilled coffee on the tablecloth.
(8) The girl was writing letters when her friend spilled coffee on the paper.

There appears to be something paradoxical about (6) in its relation to (5), which is not found in the pair (7)-(8). Whereas it belongs to the meaning of the accomplishment 'writing a letter' that the writing activity is directed toward the goal state of a complete letter, the actual occurrence of that consequent state can be denied without contradiction. How can an essential component of the meaning be denied without destroying meaning itself? This is the so-called 'imperfective paradox'. We now turn to its processing consequences.

1.3. Minimal models, inference in the limit, recomputation. Language processing amounts to incrementally computing a discourse representation given lexical, syntactic and contextual constraints [**74**]. To make computation tractable, discourse models are assumed to be 'minimal' – in a definite mathematical sense, the simplest possible structures making a narrative true. Minimal models behave like 'closed worlds': only the propositions which are asserted in discourse, or which can be inferred from it or from background knowledge, are represented as true in the model. For the remaining cases, a distinction must be drawn. The propositions that are mentioned in discourse, but are not asserted and do not follow from what is said or from background knowledge (e.g. the antecedent of a conditional), are represented as false in the minimal model. In logical terms, these propositions still belong to the finite language upon which the construction of the minimal model is based. But as long as nothing forces their truth, they will be taken as false. The propositions that are not part of the finite language – because they do not occur in the discourse context or in background knowledge – are not included in the minimal model, that is, they are not represented as true or false.

One important upshot of the theory is that the occurrence of a goal state can be inferred from a minimal model of a discourse containing an accomplishment in the past progressive. As soon as the sentence 'The girl was writing a letter' is processed, the system constructs a minimal model in which the goal state (a complete letter) is attained at some time later than the interval referred to by the progressive. Two remarks concerning this crucial point are in order. First, interpretation is based on the 'closed world assumption': if no disabling condition is described in discourse (so far), it will be (temporarily) assumed that there is no obstacle interfering with the writing activity. Second, the conclusion that eventually a letter is accomplished is an instance of *predictive inference* or, more precisely, *inference in the limit*: given that writing is asserted to hold some time in the past, that it can be assumed there are no obstacles for the writing activity, that some form of inertia holds (writing continues if it is not hindered by external forces), and that a letter is a finite object, it can be expected that the process will converge – 'in the limit' – to a complete letter. The formal definition of a 'completion inference' given in chapter 2 captures this idea in a rigorous manner. This inference is drawn for both neutral (5) and disabled (6) accomplishments. If, however, the initial model is extended with a 'when' clause describing an event which terminates the writing activity (that is, a disabling condition), the completion inference is suppressed. The subordinate clause 'when her friend spilled coffee on the paper' will lead to the retrieval of causal knowledge from semantic memory to the effect that the coffee accident terminated the writing activity. Spilling occurred *during* the writing process, from which follows that the accident took place *before* a complete letter was obtained. Here the writing event can be imagined as an open interval, where the goal state (a complete letter) is no longer part of the structure. We will use the term 'recomputation' to refer to the suppression of the goal state inference when the subordinate clause in (6) is processed. Because (5) describes a neutral scenario, the goal state derived while processing the progressivized VP is maintained in the final model. In conclusion, whereas (5) involves an *extension* of the initial discourse model, (6) might induce a *recomputation*. Because (7) and (8) do not involve a canonical goal, they will require an extension only.

1.4. Predictions for ERPs. The only difference between neutral and disabled *activities* (e.g. 'writing letters') is the noun in the subordinate clause, 'tablecloth' or 'paper'. In both cases the initial model is simply extended, therefore we expect to observe only a local ERP difference related to the integration of the differing nouns. 'Tablecloth' is less semantically expected in the context of the other lexical items occurring in the sentence compared to 'paper', so we expect a larger N400 for the former compared to the latter word.

Processing a ‘when’ clause following an *accomplishment* (e.g. ‘writing a letter’) involves integrating the differing nouns *and*, in the disabling case, recomputing the initial discourse representation. Also in this case, the neutral noun ‘tablecloth’ is predicted to evoke a larger N400 compared to the disabling ‘paper’, reflecting a lower degree of semantic relatedness with the preceding context. In our ERP study, Dutch materials were used where the verb in subordinate clauses occupies the sentence-final position (see 2.1). The temporal and causal information provided by verbs in ‘when’ clauses is necessary to initiate the recomputation process. Thus, the ERP effects of what we have analyzed as recomputation are expected to surface at the sentence-final verb ‘spilled’ (‘morste’ in the Dutch case; recall chapter 2, and see below).

One additional prediction is that the amplitude of the ERP effect evoked by disabled accomplishments is correlated with the relative frequency with which readers infer that the goal state was not attained. Recomputation is expected to consistently evoke a time-locked shift in the EEG in every trial in which a negative judgment concerning the attainment of the goal is made. Therefore, the higher the frequency of such inferences – that is, the larger the number of trials in which recomputation occurred – the larger the amplitude of the ERP component. The method and results of an ERP study in which these predictions were tested are described below.

2. Method

2.1. Materials. The Dutch stimuli used in the experiment comprised 160 test and 160 filler items. The set of materials was the same as that used in the behavioral study of chapter 2. Test items included two context sentences (O) providing a neutral setting for the events, four target sentences (A)-(D) and two probe pairs (E)-(F) (see table 1). Target sentences were constructed varying the aspectual class of the progressive VP (activity or accomplishment) and the effects of the event introduced by the ‘when’ clause (neutral or disabling) on the event described in the main progressive clause. All VPs in the progressive were instances of the Dutch periphrastic ‘was/waren NP aan het V_{inf}’ construction. This solution is to be preferred to the use of the Dutch simple past which, in some cases, is aspectually ambiguous between perfective and imperfective readings. Accomplishments differed from activities in the object NP only: an indefinite (‘een brief’/‘a letter’) was used for accomplishments, and a bare plural (‘brieven’/‘letters’) was used for activities. Disabling and neutral subordinate clauses differed only in the prepositional or object NP, for instance ‘papier’ and ‘tafelkleed’. Events were classified as neutral or disabling based on the experimenters’ judgment (see however chapter 2 and section 2.2.2 for some data supporting these choices). Probe pairs (E) were used with activities and (F) with accomplishments (table 1).

Fillers were 160 sentences of varying length, structure and content. Analogously to test items, fillers were preceded by two neutral context sentences and followed by a probe pair. Target sentences described an event consistently, as in (9), or inconsistently, as in (10), with factual knowledge (see Hagoort et al. [**79**] for an experiment based on these stimuli):

(9) Dutch trains are white and very crowded.
(10) Dutch trains are yellow and very crowded.

Probes were of the type ‘Trains in the Netherlands are white.’/‘Trains in the Netherlands are yellow.’ These fillers were chosen to add some variety to the materials while using the same task as that of test items.

Four test versions were constructed. Each of these was a randomized lists of test and filler items. The task was identical for critical and filler sentences. Participants had to select the correct probe based on the information provided by the context and target items. Mean length, raw and lemma frequency of the differing nouns in the NP of subordinate clauses were matched using the CELEX Dutch corpus [**4**]. Mean length was 7.9 letters (SD=2.46) for neutral and 7.75 (SD=2.79) for disabling nouns, and was kept below 12 letters in any case. Raw frequency per million words was 1113 (SD=2462) for neutral and 1096 (SD=2792) for disabling nouns. Lemma frequency per million words was 1730 (SD=3559) for neutral and 1666 (SD=3585) for disabling nouns. The length of sentence-final verbs was identical across conditions and was kept below 12 letters in any case.

	Context sentences
(O)	De deur van de woonkamer was gesloten. Binnen speelde de radio klassieke muziek. *The door of the living-room was closed. Inside played the radio classical music.*[a] 'The door of the living room was closed. Inside the radio played classical music.'[b]
	Target sentences
(A)	Het meisje was brieven aan het schrijven toen haar vriendin koffie op het tafelkleed morste. *The girl was letters on the to-write when her friend coffee on the tablecloth spilled.*[a] 'The girl was writing letters when her friend spilled coffee on the tablecloth.'[b]
(B)	Het meisje was brieven aan het schrijven toen haar vriendin koffie op het papier morste. *The girl was letters on the to-write when her friend coffee on the paper spilled.*[a] 'The girl was writing letters when her friend spilled coffee on the paper.'[b]
(C)	Het meisje was een brief aan het schrijven toen haar vriendin koffie op het tafelkleed morste. *The girl was a letter on the to-write when her friend coffee on the tablecloth spilled.*[a] 'The girl was writing a letter when her friend spilled coffee on the tablecloth.'[b]
(D)	Het meisje was een brief aan het schrijven toen haar vriendin koffie op het papier morste. *The girl was a letter on the to-write when her friend coffee on the paper spilled.*[a] 'The girl was writing a letter when her friend spilled coffee on the paper.'[b]
	Probe sentences
(E)	Het meisje heeft een of meer brieven geschreven. *The girl has one or more letters written.*[a] 'The girl has written one or more letters.'[b] Het meisje heeft geen brief geschreven. *The girl has no letter written.*[a] 'The girl has written no letter.'[b]
(F)	Het meisje heeft een brief geschreven. *The girl has a letter written.*[a] 'The girl has written a letter.'[b] Het meisje heeft geen brief geschreven. *The girl has no letter written.*[a] 'The girl has written no letter.'[b]

TABLE 1. Examples of stimulus sentences. [a]Literal translation. [b]Paraphrase.

2.2. Pre-tests.

2.2.1. *Cloze probability test.* To determine the cloze probabilities of sentence-final verbs, context sentences followed by a target sentence with the final word blanked were presented to a group of thirty-two native speakers of Dutch (mean age 22.5, 27 female). Participants were requested to fill in the blank with the first word that came to their mind. Four versions (40 items per condition), randomized and balanced across conditions, were constructed. Mean cloze probabilities were not different between the conditions (all comparisons using T-tests, $P > 0.05$) in each test version as well as in the entire set.

2.2.2. *Entailment questionnaire.* A paper-and-pencil judgment task was administered. Thirty six Dutch native speakers (mean age 22.5, 28 female) were presented with the context followed by a target sentence and a probe pair. The task was to select the appropriate probe. Negative probes were more frequently chosen for disabled accomplishments than for the other conditions. Neutral activities (A) (see table 1) showed the lowest mean of negative responses (M=2.72, SD=3.22), followed by disabled activities (B) (M=8.06, SD=7.05), neutral accomplishments (C) (M=10.03, SD=9.23) and disabled accomplishments (D) (M=25.14, SD=8.02) (see 2 for details).

2.3. Participants. Thirty one students participated in the ERP experiment. Of these, 7 were left out of the final analysis due to a high number ($>$ 20%) of trials contaminated by artifacts. This left us with twenty four participants (mean age 22.5, 17 female), with no history of neurological, psychiatric or cognitive disorders. Subjects were selected from the database of the Donders Centre for Cognitive Neuroimaging at the Radboud University Nijmegen. Participants received € 8 per hour or course credits. None of the subjects who took part to the pre-tests and to the behavioral study reported in chapter 2 participated in the ERP experiment.

2.4. Procedure. After applying the electrodes, participants were conducted into the experimental booth and were asked to sit in front of the video monitor. The stimuli were presented as follows: the two context sentences were displayed together on a single screen (white on black background) for a variable duration (6, 7 or 8 s), depending on the length of the sentences themselves; next the target sentence, one of (A)-(D), was presented on the screen word-by-word (600 ms SOA, 300 ms word duration; white on black background); the target sentence was preceded and followed by a fixation cross, presented for 1500 ms; finally, the probe pair, one of (E)-(F), was shown on the screen (red on black background) and remained visible until the participant gave a button-press; the probes were followed by a fixation cross which lasted for 1500 ms. The same presentation parameters were used for fillers. Participants were instructed to read each sentence carefully, to blink only when the fixation cross was shown, and to select the correct probe by pressing one of two buttons (left or right on the button box) as quickly and accurately as possible. The position on the screen (top or bottom) of the positive and negative probe corresponded to the left and right button respectively, and was counterbalanced across test versions. In this way, participants could not prepare their motor response before the probe pair was presented on the screen. The experiment took about 2 hours and was divided into 24 blocks of 10 trials each.

2.5. Recording. EEG and EOG signals were recorded using Ag/AgCl electrodes. The EOG was measured from 4 electrodes: one at the outer canthus of each eye, one below and one above the left eye (FE). The EEG was measured from 28 electrodes, arranged according to American Electrophysiological Society conventions: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CE, CF, CP6, P7, P3, Pz, P4, P8, O1, O2. Two additional electrodes were placed on the left and right mastoids, the former serving as the reference during the measurement. All EEG and EOG electrodes were re-referenced off-line to a linked mastoid. EEG electrodes were attached to an elastic cap, whereas EOG and reference electrodes were applied using two-sided adhesive decals external to the cap. Electrode impedance was kept below 5 kΩ throughout the experiment. The EEG/EOG was amplified by a multichannel BrainAmp DC system, with a 500 Hz sampling rate, a low pass filter set at 125 Hz and a 10 s time constant.

2.6. Data Analysis. For the data analysis we used FieldTrip,[1] a MATLAB package for processing EEG and MEG signals. The following transforms were applied to each subject's dataset. Segments corresponding to the critical noun and to the sentence-final verb were extracted from the EEG with an interval of 200 ms preceding and 800 ms following word

[1]For more information, see `http://fieldtrip.fcdonders.nl/`

onset. Baseline correction used the 200 ms interval preceding the onset of the critical noun, and the 100 ms interval following the onset of the sentence-final verb. The latter choice was effected so as to prevent ERP differences in the 400-600 ms interval following the onset of nouns from biasing the baseline correction for the ERPs evoked by sentence-final verbs in the same temporal interval. The use of such a post-stimulus baseline seems acceptable on grounds that the expected recomputation effect at the verb should not affect such largely exogenous components as the N1. Artifact rejection was based on two FieldTrip functions: the first detects and rejects all trials that contain activity exceeding a threshold of $\pm 100\ \mu V$; the second identifies and discards trials contaminated with eye movements or blinks by means of thresholding the z-transformed value of the raw data in the EOG channels, pre-processed using a band-pass filter of 1-15 Hz. A 30 Hz low-pass digital filter was applied to the segmented, artifact-free data. ERPs were obtained for each subject by averaging over trials in each experimental condition. A 5 Hz low-pass filter was used to produce the wave-forms shown in figures 2-5. Topographies and statistical analyses were based on the 30 Hz low-pass filtered data.

For the analysis of behavioral responses we employed two repeated-measures ANOVA models with Subject as the random effect, Aspectual Class (Activity/Accomplishment) and Subordinate Clause Type (Neutral/Disabling) as fixed effects, and the mean value of either negative judgments (negative probes selected in the response task) or decision times in each condition as the dependent variables.

Statistical analyses of ERP data used a non-parametric randomization procedure [**129, 130**] which was based on mean amplitude (μV) values in each condition in time bins of 100 ms, starting from the onset of the relevant word and ending 800 ms after. The algorithm produced as output a cluster of electrodes (min. 1, max. 28) in which the difference between the conditions was significant in each time bin, the sum of T-statistics in that cluster and Monte Carlo estimates of P-values.

For the correlation analysis (see section 1.4), we calculated the difference between the ERPs evoked by sentence-final verbs in subordinate clauses – disabled (D) minus neutral (C) – following accomplishments at anterior sites (Fp1, Fp2, F7, F3, Fz, F4, F8 averaged) in the 500-700 ms interval after the onset of the sentence-final verb (see 3.2 for motivation). Pearson's product-moment correlation was computed to determine whether the amplitude difference in ERPs varied with the number of negative responses, quantified again as the difference of negative judgments between disabled (D) and neutral (C) accomplishments. The correlation analysis was done on a per-subject basis, so that each pair of data points in the correlation corresponded to a single subject's data.

3. Results

3.1. Behavioral data. Neutral activities (A) had the lowest mean of negative responses overall (M=4.08, SD=2.87), and were followed by disabled activities (B) (M=5.83, SD=4.51), neutral accomplishments (C) (M=9.58, SD=9.96) and finally disabled accomplishments (D) (M=18.13, SD=11.16). The distribution of the data in the different conditions appears rather similar. Interestingly however, as can be seen in figure 1-a, disabled accomplishments have a more spread-out distribution, suggesting that inference patterns were less uniform across participants. ANOVAs revealed significant main effects of Aspectual Class and Subordinate Clause Type, and a significant interaction between the two factors (table 2, figure 1-a). The observed pattern of responses supports the linguistic views outlined above and replicates our previous findings (see section 2.2.2 and chapter 2). We found no difference in decision times (table 2, figure 1-b): (A), M=2111 ms, SD=677 ms; (B), M=2100 ms, SD=680 ms; (C), M=2070 ms, SD=646 ms; (D), M=2086 ms, SD=710 ms.

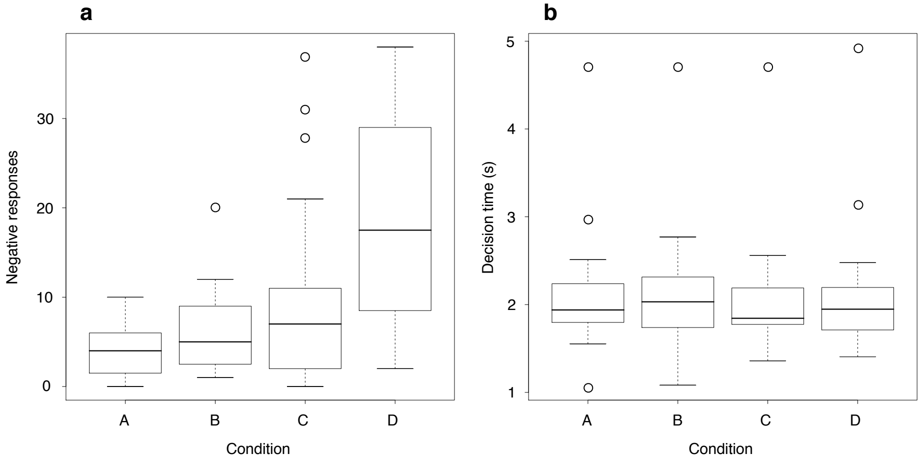


FIGURE 1. Behavioral data. (a) Boxplot of the categorical responses. (b) Boxplot of the decision times. Conditions are represented on the abscissa (see table 1 for the correspondences with the labels). Negative responses and decision times are plotted on the ordinate. The solid line within the boxes indicates the median, box height is equal to the interquartile range, whiskers represent adjacent values, and empty circles denote outliers. The maximum of potential negative responses is 40.

3.2. Event-related brain potentials.

3.2.1. *Nouns.* Figure 2 displays the ERP topographies and waveforms elicited by *activities*. An N1 component peaking at approximately 100 ms is followed by a P2 component with a trough at about 200 ms. The amplitudes of the N1 and P2 are not different between disabled and neutral clauses: no significant clusters were found between 0 and 300 ms (all contrasts, $P > 0.1$). The N1-P2 complex is followed by an N400. The amplitude of the N400 is larger in neutral ('tafelkleed') than in disabling ('papier') clauses (figure 2-b): significant clusters with a central distribution were found between 300 and 500 ms (table 3, figure 2-a). No difference between neutral and disabling clauses was found after 500 ms.

	Categorical responses	Decision times
Aspectual Class	$F(1,23)=21.65$ $P<0.001$	$F(1,23)<1$
Subordinate Clause Type	$F(1,23)=23.60$ $P<0.001$	$F(1,23)<1$
Aspectual Class $\times$ Subordinate Clause Type	$F(1,23)=17.20$ $P<0.001$	$F(1,23)<1$

TABLE 2. Summary of ANOVA statistics for behavioral data.

Time	Noun		Sentence-final verb	
	Activities	Accomplishments	Activities	Accomplishments
300-400 ms	$T(22) = -16.54$ $P = 0.026$	$T(22) = -60.11$ $P < 0.001$		
400-500 ms	$T(22) = -57.02$ $P < 0.001$	$T(22) = -78.69$ $P = 0.002$		$T(22) = -11.85$ $P = 0.034$
500-600 ms		$T(22) = -18.58$ $P = 0.022$		$T(22) = -71.09$ $P < 0.001$
600-700 ms				$T(22) = -39.16$ $P = 0.008$
700-800 ms				$T(22) = -16.92$ $P = 0.028$

TABLE 3. Summary of cluster-based statistics for the ERP data. Disabling and neutral clauses are compared at the noun and at the sentence-final verb, for activities and accomplishments, in bins of 100 ms starting from word onset. The first significant effects occurred at 300-400 ms. Empty cells denote the absence of significant clusters.

figure 3 displays the ERP topographies and waveforms elicited by *accomplishments*. Also in this case, an N1-P2 complex can be observed. There is no difference between neutral and disabling clauses, as no significant clusters between 0 and 300 ms were found (all contrasts, $P > 0.1$). The amplitude of the N400 is again larger in neutral ('tafelkleed') than in disabling ('papier') clauses (figure 3-b). The effect lasts longer than the N400 observed in activities: significant clusters with a central distribution were found between 300 and 600 ms (table 3, figure 3-a). For activities, no difference between conditions was found after 500 ms.

There is no overall difference between the two aspectual classes. Cluster-based T-tests comparing the N400 effects in activities and accomplishments (corresponding to testing the main effect of Aspectual Class in a parametric model) produced no significant clusters between 300 and 600 ms from noun onset (all contrasts, $P > 0.1$). No difference was found in any of the remaining time bins.

3.2.2. *Sentence-final verbs.* Figure 4 shows the ERP topographies and waveforms elicited by *activities*. Contrary to what we had observed at the noun, there is no difference between the N400 elicited by neutral and disabling clauses. Moreover, there is no difference between conditions in any of the remaining time bins (table 3).

Figure 5 displays the ERP topographies and waveforms elicited by *accomplishments*. No difference between disabling and neutral clauses was observed in either the N1-P2 complex or the N400: no significant clusters between 0 and 400 ms were found (all contrasts, $P > 0.1$). Whereas disabled and neutral *activities* do not result in any robust differential effect in later time bins (400-800 ms, table 3, figure 4), disabling verbs following *accomplishments* evoked larger negative shifts compared to neutral verbs (table 3, figure 5). This effect emerges at about 400 ms following the onset of sentence-final verb, lasts for approximately 400 ms, and is larger over left-anterior scalp sites. Based on its temporal profile and scalp distribution, we take this effect to be an instance of sustained anterior negativity (SAN). The magnitude of the SAN effect is correlated with the frequency of negative judgments in the response task ($r = -0.415, T(22) = -2.140, P = 0.043$; figure 6): the higher the number of negative responses, the larger the amplitude of the SAN.

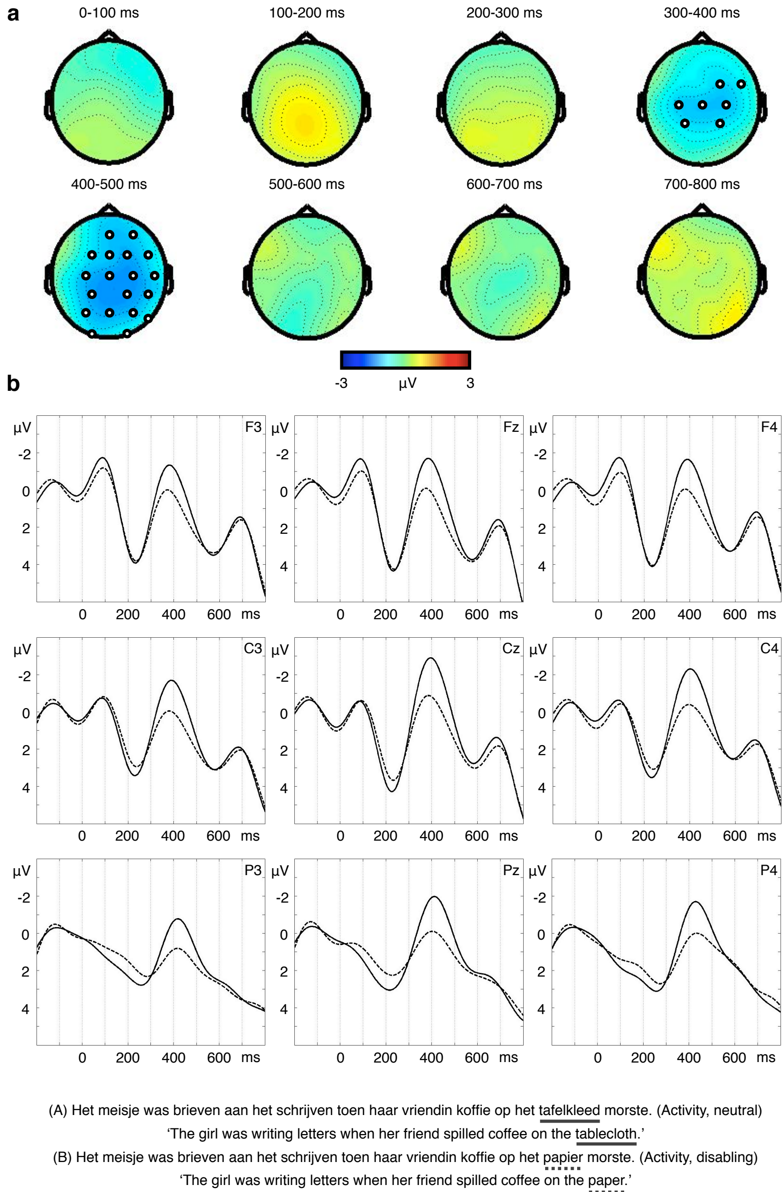


FIGURE 2. Activities, noun. (a) Grand-average (N=24) topographies displaying the mean amplitude difference between the ERPs evoked by the noun in neutral compared to disabled activities. Circles represent electrodes in a significant cluster. (b) Grand-average (N=24) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the noun in neutral and disabled activities. Negative values are plotted upward.

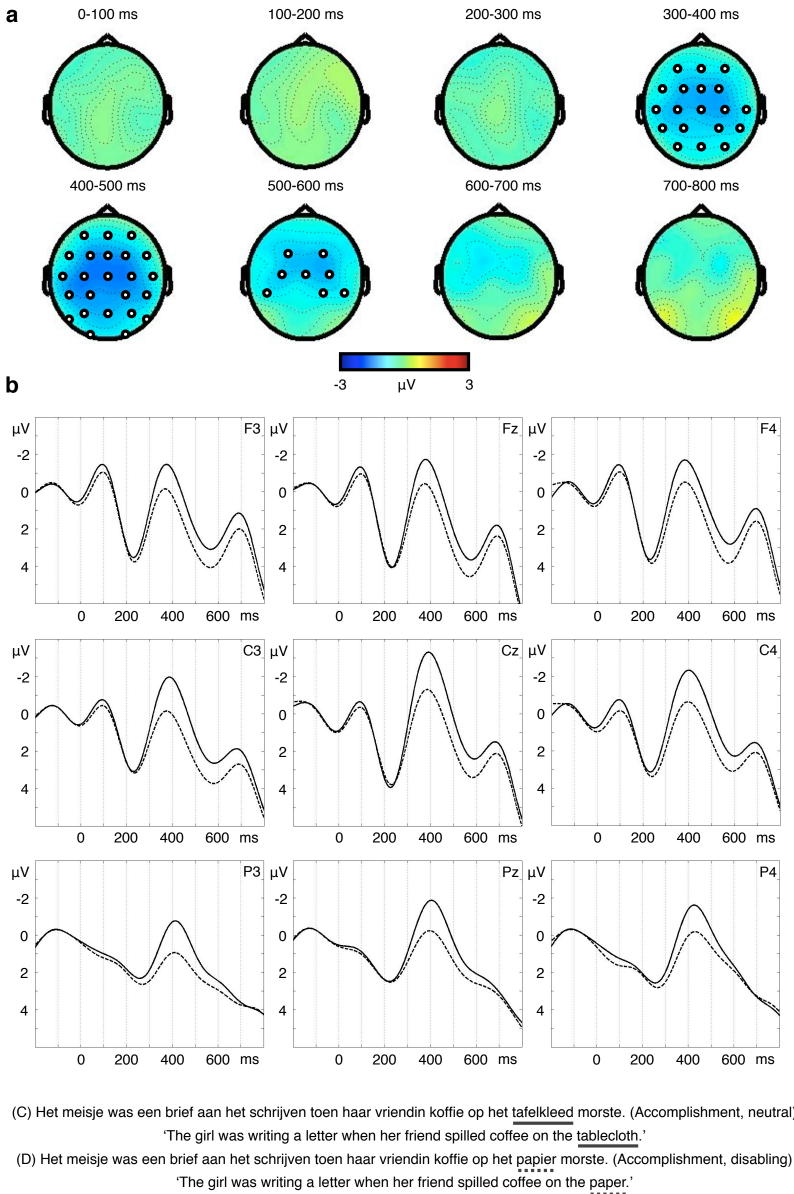


FIGURE 3. Accomplishments, noun. (a) Grand-average (N=24) topographies displaying the mean amplitude difference between the ERPs evoked by the noun in neutral compared to disabled accomplishments. Circles represent electrodes in a significant cluster. (b) Grand-average (N=24) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the noun in neutral and disabled accomplishments. Negative values are plotted upward.

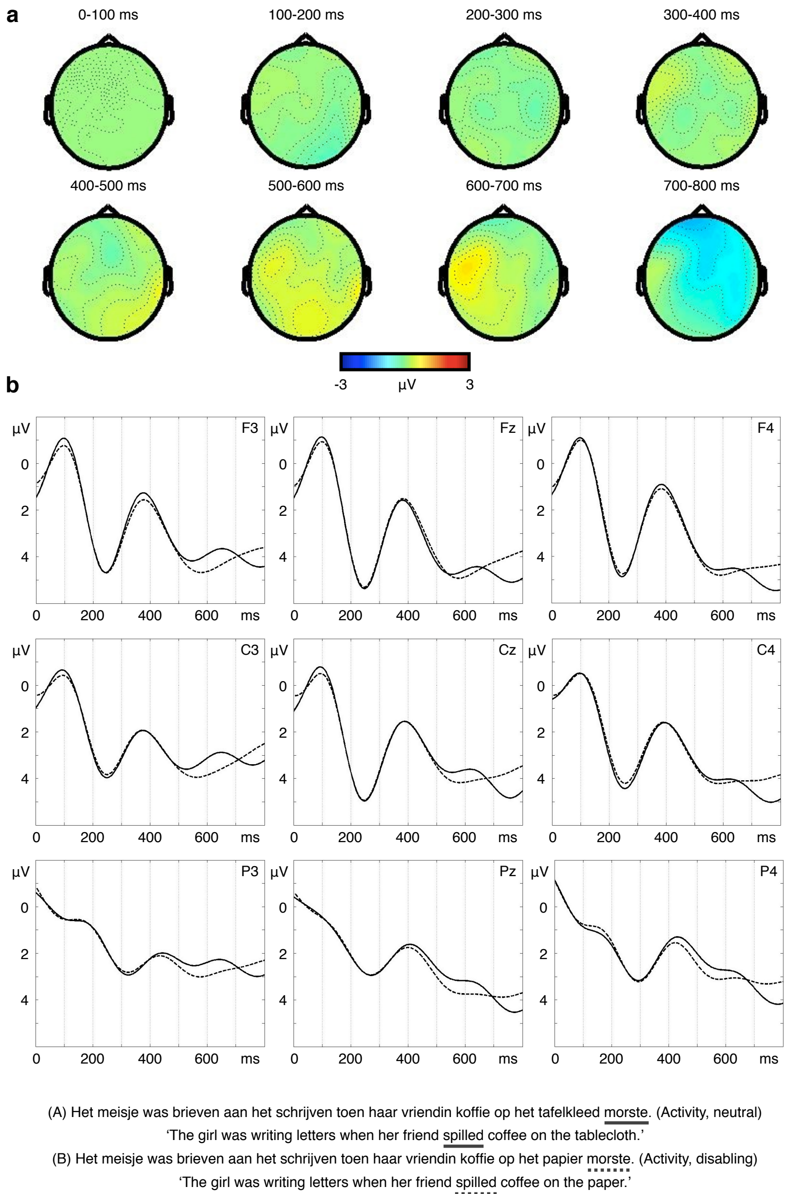


FIGURE 4. Activities, sentence-final verb. (a) Grand-average (N=24) topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final verb in disabled compared to neutral activities. (b) Grand-average (N=24) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the verb in neutral and disabled activities. Negative values are plotted upward.

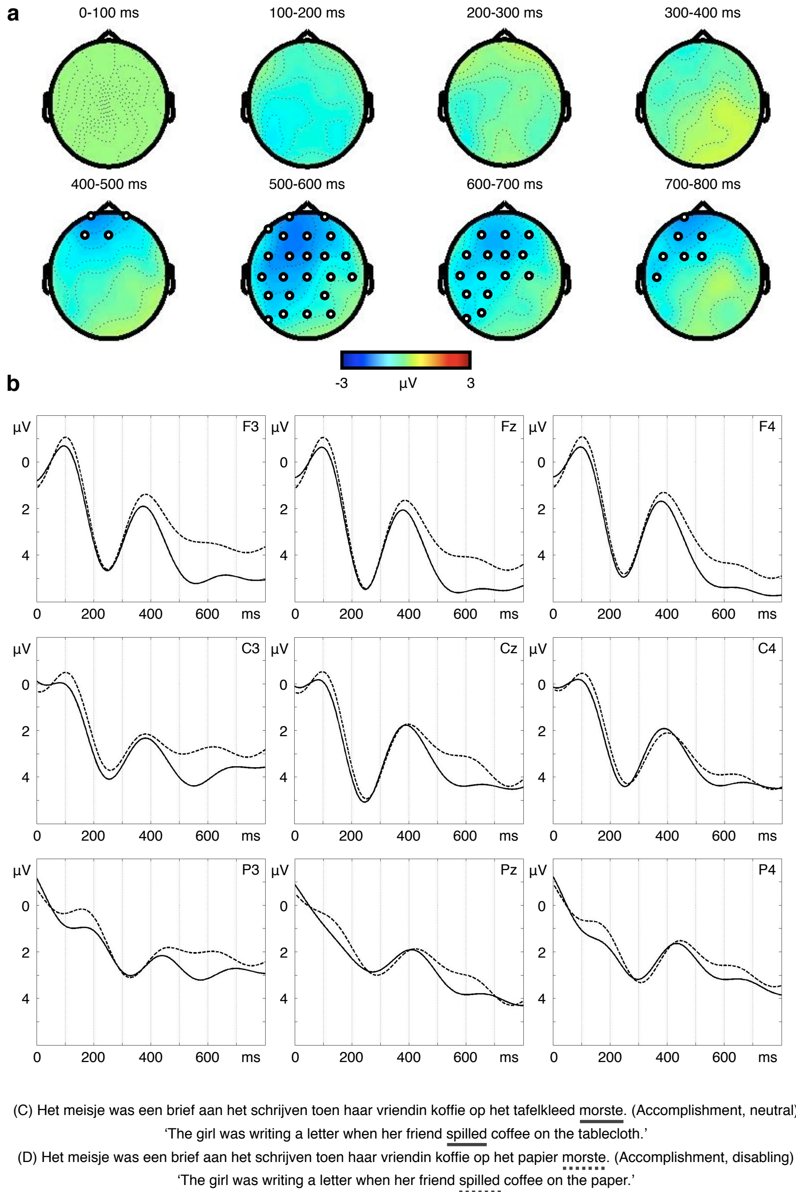


FIGURE 5. Accomplishments, sentence-final verb. (a) Grand-average (N=24) topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final verb in disabled compared to neutral accomplishments. Circles represent electrodes in a significant cluster. (b) Grand-average (N=24) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the verb in neutral and disabled accomplishments. Negative values are plotted upward.

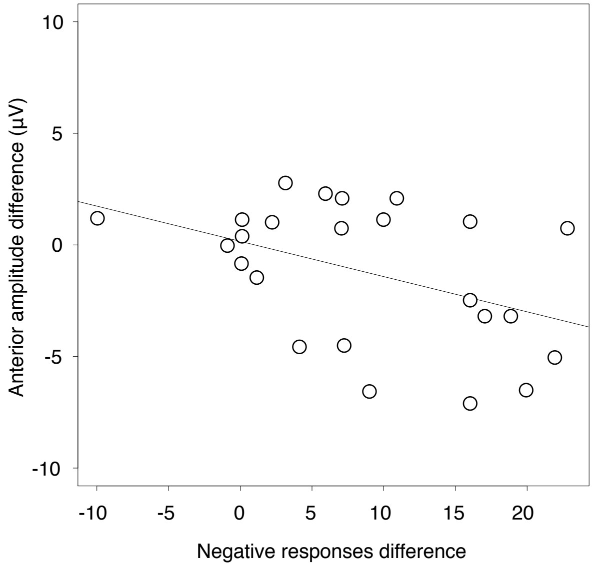


FIGURE 6. Scatter plot displaying the correlation between the amplitude of the sustained anterior negativity elicited by disabled accomplishments and the frequency of negative responses ($r = -0.415, T(22) = -2.140, P = 0.043$). The mean difference of negative responses between disabled and neutral accomplishments is plotted on the abscissa. The mean amplitude difference at fronto-polar and frontal electrodes between disabled and neutral accomplishments in the 500-700 ms interval following the onset of the sentence-final verb is plotted on the ordinate.

No difference between the two aspectual classes was found. Cluster-based T-statistics comparing mean ERP amplitudes in activities and accomplishments, again corresponding to testing the main effect of Aspectual Class in a parametric model, produced no significant clusters between 0 and 800 ms from noun onset (all contrasts, $P > 0.1$).

4. Discussion

The ERP results reported above can be summarized as follows. The N400 elicited by nouns is larger in neutral than in disabling clauses, following both activities and accomplishments. This can be explained by the lower degree of semantic association with the preceding words ('writing' and 'letter' or 'letters') of the noun in a neutral clause ('tablecloth') as compared to the noun in a disabling clause ('paper'). On the basis of our processing model, we predicted that disabled accomplishments would induce a different ERP response at the sentence-final verb compared to neutral accomplishments. This corresponds to the difference between the recomputation and the extension of the initial discourse model (see section 1.3). The effect was expected to be (i) absent in activities and (ii) correlated with the frequency with which participants inferred that the goal state was not attained. These predictions were borne out. Disabled activities did not modulate ERPs at the verb. Disabled accomplishments evoked sustained anterior negativities (SANs). Moreover, a correlation of the SAN amplitude with the frequency of negative judgments was observed. Taken together, our results would seem to offer some support for the recomputation hypothesis. Below we address a few alternative explanations of the data and some related outstanding issues.

4.1. Alternative explanations and outstanding issues.

4.1.1. *Local integration.* One possible alternative explanation relates the observed effect to difficulty in integrating the sentence-final verb into the 'local', clause-level context, rather than to suppressing a 'global', discourse-level inference. If this were correct, a modulation of the N400 should be expected, correlated with differences in cloze probabilities. However, as reported above, cloze probabilities are not different between conditions (see 2.2.1). Also, there was no difference in the N400s elicited by sentence-final verbs (see 3.2), which were lexically identical across conditions. Following earlier research [**149**], we take the sustained anterior negativity as reflecting difficulty in constructing a discourse-level representation of disabled accomplishments. Supported by further evidence, recomputation could provide a more explicit characterization of at least one instance of sentence-final 'wrap-up effects', in terms of restructuring an existing sentence or discourse model.

4.1.2. *Response frequency.* Another alternative account is based on the observation that sentences requiring a negative response (disabled accomplishments) are less frequent than sentences requiring a positive one (activities and neutral accomplishments), the projected ratios being respectively 1/4 and 3/4 (see 3.1 for the actual behavioral data). On this view, a modulation of the P300 component [**40, 181**] might be expected, inversely correlated with the frequency of negative judgments given to disabled accomplishments: the less frequent the negative responses, the larger the amplitude of the P300. However, in our data no P300 response was observed, and the correlation was rather the inverse: the more frequent the negative responses, the larger the amplitude of the sustained anterior negativity.

4.1.3. *Monotonicity and possible worlds semantics.* A crucial issue is whether the observed sustained anterior negativity can be explained by a monotonic account of the progressive. That is, it may be asked whether the ERP data reported here constitute compelling evidence for non-monotonicity and against monotonicity. One such alternative explanation can be formulated in possible worlds semantics [**114, 42**].

As we saw in chapter 2, in possible worlds semantics the progressive denotes a stage of a process which, if it does not continue in the actual world, has chances of continuing in some other possible worlds [**36**]. These can be called 'inertia worlds', courses of events in which the process is not disturbed by external forces and is therefore brought to a successful end. In his analysis of the progressive, Dowty [**42**] claimed that the following are equivalent:

1. 'The girl is writing a letter' is true in the actual world;
2. 'The girl will have written a letter' is true in all so-called 'inertia worlds', worlds which are identical with the present world until 'now', but then continue in a way most compatible with the history of the world until 'now'.

This view has the following processing consequences. Processing neutral accomplishments involves moving from the actual world, a snapshot of which is provided by the progressive clause, to some inertia world, in which the goal state is eventually attained (the behavioral data reported in 3.1 show that such an inference is drawn). By contrast, processing disabled accomplishments amounts to proceeding from the actual world to a relatively unexpected 'non-inertia world', in which the process is disrupted by some event, such as spilling coffee on the paper. Accessing a world in which the goal is not attained may be surprising. The sustained anterior negativity may then be construed as an index of surprise or an equivalent notion. This account is monotonic, as both neutral and disabled accomplishments require shifting from the actual world to another accessible world. A simple extension of the initial model is performed in both cases.

This analysis is seemingly in conflict with the non-monotonic one. Nevertheless, there seems to be no real opposition between the two. In both accounts, an initial commitment to the occurrence of the goal state is made. In the non-monotonic approach this takes the form of a defeasible inference licensed by the minimal model, whereas in the monotonic theory it is rather a prior, positive expectation concerning the attainment of the consequent state, implying a lower probability assigned to its failure. This commitment is necessary, because accessing a possible world in which the goal state is not satisfied can be surprising only if there is such a prior expectation. Furthermore, in both accounts a subsequent revision of the initial commitment is made. In the non-monotonic analysis this is a recomputation of the initial minimal model, whereas in the monotonic one it is a recomputation of the initial (low) probability associated with the possible world in which the goal state is not attained.

There is recomputation and non-monotonicity is built into this 'monotonic' theory too. The possible worlds account is monotonic as far as *models* are concerned (models are always extended and never recomputed), but *expectations* change non-monotonically (probabilities are recomputed). While *prima facie* opposed, the two accounts are in this respect equivalent. Our reasons for preferring a non-monotonic account, one in which *models* are recomputed, derive from theoretical considerations. First, there exist forms of non-monotonic inference formally strongly related the non-monotonic reasoning in the progressive which cannot be captured by Bayesian updates [**196**]. Second, minimal models can be computed efficiently and non-monotonic inference can be implemented in neural networks (see 4.2), whereas it is unclear to what extent models involving possible worlds and an accessibility relation are at all computable. Last, the minimal models account of the progressive is embedded into a larger non-monotonic framework [**212, 195**], covering other phenomena in reasoning and language processing in children, adults, as well as patients with ADHD (Attention Deficit Hyperactivity Disorder) [**214**] and ASD (Autistic Spectrum Disorder) [**213**].

We must note however that our study was designed to test a particular non-monotonic theory of the progressive, and not to discriminate between monotonic and non-monotonic accounts of the same phenomenon. The latter task would require, for one, a well-specified *entirely* monotonic theory – that is, one that does not involve the recomputation of models, probability values or other processing parameters – and, moreover, a set of predictions in which the two proposals would actually differ. This is admittedly hard, apart from being beyond the scope of the research reported above. Hence the need to emphasize the direction along the theory-observation path which is relevant here: although it can be argued that our non-monotonic theory leads to predictions that are consistent with the observed sustained anterior negativity, it is clearly not the case that the data support *only* this particular theory.

4.1.4. *Interruption and termination of activities.* Compared to disabled accomplishments, disabled activities may be inherently simpler because they involve at most an *interruption* of the activity, for example writing letters in (B), which might be continued on some other paper sheets. Accomplishments might leave a more definite ERP trace because they lead to the *termination* of the activity, for example writing a letter in (D), which cannot be continued being there only a single sheet. On this account, the sustained anterior negativity is not due to model recomputation (as opposed to monotonic extension), but to the termination (as opposed to the interruption) of the activity. Such an explanation follows from the seemingly plausible notion that computing a discourse model where the effects of an event are more 'catastrophic' should also be more difficult. Here semantic theory comes to our rescue and suggests that such notion is in fact misguided.

One issue that plays a role here is a type/token distinction concerning the noun 'letter'. In the token interpretation of 'letter' as referring to some particular scribbles on a particular piece of paper, there is indeed a difference between interruption and termination. However, on a type interpretation of 'letter' as referring to particular content which can be inscribed on any piece of paper, the activity and the accomplishment case seem comparable, in that in both cases the girl has to reach for a new piece of paper. On the type reading, one wouldn't even expect a difference in behavioral responses. Nonetheless, since a behavioral difference was observed, it seems the token reading is what subjects adopt. On this assumption, it can be shown that, contrary to the alternative proposal, there is *more* computation going on in the interruption case compared to the termination case – if goal states are not taken into account; if they are, the pattern is reversed as implied by the recomputation hypothesis. It seems harder to compute a model in which an activity is first interrupted, then re-initiated, compared to computing a model in which the activity is just terminated [**212**]. The alternative account would predict a larger sustained anterior negativity for activities compared to accomplishments, which does not fit the experimental results. Also in this case, however, we are ready to acknowledge that a different model, in which terminations are shown to be *more* costly than interruptions, and in which goal states are not invoked to account for such processing cost, may explain the observed sustained anterior negativity.

4.1.5. *Goal states and underspecification.* Our processing theory implies that, when an accomplishment in the past progressive is encountered, the system computes a model in which the goal state is satisfied. Processing the clause 'The girl was writing a letter' amounts to computing a minimal model in which the writing activity will lead to a complete letter, which is therefore part of the resulting discourse structure. We have hypothesized that this computation is defeasible, so the model can be recomputed if further discourse information implies that the goal state is not satisfiable, as in (6). One may ask whether the claim that the culmination and consequent state are part of a minimal model of a progressive clause is at all tenable. A seemingly more plausible account would assume that an *underspecified model*, in which it is left undecided whether the goal state is attained or not, is computed while the progressive is processed, and a decision is made only at the subordinate clause.

The main problem with an underspecification-based account is that, while it is true that the information provided by the progressive clause is insufficient for determining whether the goal was attained (which would motivate the construction of an underspecified model at that stage), it is *not* the case that sufficient information is contributed by the subordinate clause. While disabling clauses provide evidence that the activity was terminated, and thus license the inference that the goal was not attained, no evidence concerning the satisfaction of the goal state is derivable from neutral clauses. This is a consequence of the well-known 'frame problem' [**134**], which implies that it is impossible to enumerate all the effects and non-effects of an event. For example, that 'spilling coffee on the tablecloth does not affect the writing activity' (if that is the case) cannot be stored in declarative memory, but must be inferred. This is an instance of 'closed world reasoning', which we described earlier on. In a 'closed world', unexpected accidents are ruled out. Therefore, the letter was completed. The behavioral data reported above show that subjects draw this inference. Processing models based on underspecification – or parallel processing, for that matter – would have to explain why that very same conclusion ('the girl wrote a letter') is not drawn when the system is faced with the relevant input (the VP in the progressive), and is instead delayed until the end of the sentence, where critical information is nonetheless still missing. The hypothesis that the goal state inference is drawn when the input is given seems to us more consistent with the available evidence on immediacy and incrementality in discourse processing [**80**].

This line of reasoning speaks also to the issue of the possible influence of the primary task on on-line comprehension processes. It can be argued that the system may have carried out a number of inferences on-line in order to facilitate the response when the probes were presented, but would have processed the same sentences in an underspecified manner if no response task was administered. The brain would therefore compute representations which are merely 'good enough' for the task at hand, striking a balance between efficiency and cost minimization [**49, 50, 41**]. We grant this as a possibility, which cannot be excluded based on either our data or our processing model. However, if it is true that comprehension probes do not occur in actual language use, it is possible to imagine 'language games' in which hearers are required to make interpretive commitments and form a belief concerning the potential outcomes of a process described using the progressive. Our experiment may be seen as a laboratory construction of such real world situations, but is not intended as a realistic account of all situations in which progressive sentences are uttered and understood. Further research is necessary to investigate the influence of the primary task on ERPs. This is a question that applies to most experimental paradigms, and would require systematic investigation.

4.2. Recomputation in working memory networks. Minimal models correspond to the stable states of associated neural networks. It has been shown that recurrent networks can compute or approximate (depending on the expressiveness of one's logical language) the semantic operators based on which minimal models can be constructed [**91, 194, 195**]. In this framework, recomputation can be modeled as the readjustment of connection strengths driven by a simple form of back-propagation called 'perceptron learning' [**179**]. Computing a minimal model of the progressive clause will correspond to the network settling into one such attractor or stable state. Further computation on the initial model brings the network from its initial stable state to another stable state, corresponding to the new minimal model. Importantly, there is a large difference in the overall pattern of network activity in disabled compared to neutral accomplishments. If the initial model is monotonically extended, as in the neutral case (4), a number of units will be activated which were previously silent, while the activation state of the remaining units, including those representing the goal state (the complete letter), will remain unaltered. But if the initial minimal model is recomputed upon encountering the subordinate clause in (6), units which were silent will be activated *and* the activation patterns across some units which were previously active will be readjusted. For instance, the units representing the goal state (the complete letter) will be shut down. In the neural network this is achieved by successive applications of perceptron learning. It is still unclear which biological mechanism could perform this function.

Even though in both cases the network processes the subordinate clause by settling into a new attractor state, the transition in the disabling case requires an extensive adjustment of the connection weights of the units representing the goal state. Recomputation thus results in a more costly state transition. It remains an open question whether biologically plausible networks can also approximate the semantic operators which give rise to minimal models. Firing rate models, for instance, have been used to implement operations in connectionist networks (e.g. multilayer perceptrons) of the kind required by the construction of minimal models [**34**]. Interestingly, recurrent excitation in firing-rate models can account for several aspects of persistent activity in prefrontal cortex neurons during working memory tasks [**43**]. Recurrent networks thus suggest a plausible mechanistic link between recomputation and sustained anterior negativities, and in general between working memory processes and slow negative potentials [**112, 146, 147, 207, 53, 48, 206, 161**].

The leitmotiv of this book is that the cognitive neuroscience of language needs to bridge the gap between psycholinguistic and formal models of specific aspects of language on the one hand, and the neural architecture underlying neurophysiological measures on the other hand. For a number of reasons [**165**] this is a daunting task, which we do not claim to have adequately solved. However, tentatively the following can be said. There is no indication or proof that the sustained anterior negativity is a language-specific ERP effect. Most likely, it reflects the recruitment of neurophysiological activity that might be generated in prefrontal cortex, and is triggered by different cognitive operations building upon working memory capacity. That is why the prefrontal cortex is a plausible candidate from a neurobiological point of view. In the light of our model, the sustained anterior negativity is taken to index the recomputation following the termination of the activity in accomplishments, and the recruitment of working memory resources required for this recomputation. In other cases, the demand might be triggered by different cognitive operations [**147**]. In general, what we seem to obtain with ERPs and other techniques is a many-to-one mapping from cognitive models to neuronal implementation. This however in no way invalidates our model, which is based on combined constraints from the cognitive and neuronal levels of analysis.

The research presented in this chapter extends the range of phenomena to which ERPs can be applied, by testing a processing hypothesis suggested by a formal semantics of tense and aspect. Our results indicate that the brain might support some form of non-monotonic recomputation to integrate information which invalidates previously held assumptions. It is a task for future research to provide more stringent tests of non-monotonic as opposed to monotonic theories of semantic processing and cognitive update more generally.

CHAPTER 6

Coercion and compositionality

This chapter is a modified version of G. Baggio, T. Choma, M. van Lambalgen & P. Hagoort. Coercion and compositionality. Under review.

1. Introduction

Productivity is the hallmark of language. Speakers of a language are capable of producing and understanding a seemingly unlimited number of novel sentences starting from a finite repository of lexical meanings and morpho-syntactic constraints. Productivity requires that the brain is endowed with some computational mechanism combining stored signals into larger, and possibly novel expressions. The principle of compositionality is usually invoked to account for the productivity of language, as did Katz and Fodor [**108**]. Compositionality is the notion that the meaning of a complex expression is a function *only* of the meanings of its constituent parts and of the way they are syntactically combined [**159**]. This view has at least one important consequence for theories of language processing: if word meanings are assembled based on syntactic constraints alone, there is no other source of combinatoriality in language than the syntax [**33**]. However, recent studies in psycholinguistics suggest that language comprehension cannot be reduced to a strict, syntax-driven composition of lexical meanings, but involves an independent level of semantic combinatorics which can, in some cases, constrain syntactic analysis [**111**], and makes use of world knowledge and discourse context – beyond context-invariant elementary meanings – to arrive at a full interpretation of the utterance [**80**]. The processing consequences of compositionality have been discussed in chapter 3. In this paper, ERP data will be presented that challenge strictly compositional views of meaning assembly.

1.1. Complement coercion and enriched semantic composition. A great deal of the experimental evidence supporting the independence of semantic operations has come from studies of complement coercion [**137, 205, 204, 136, 173**]. Verbs like 'begin' can occur in two types of syntactic constructions with similar meanings. (a) As a subject control verb with an NP and a VP arguments: 'The journalist began to write the article' or 'The journalist began writing the article'. Or (b) as a transitive verb with two NP arguments: 'The journalist began the article' [**44**]. In a number of cases, constructions of the second type involve an implicit semantic element, which is expressed in the first type by the head of the VP argument. The contrast between silent and expressed meaning is also found in the following pair:

(1) The journalist wrote the article before his coffee break.
(2) The journalist began the article before his coffee break.

In (1), the activity that the journalist is performing is expressed by the verb 'wrote', while in (2) the activity which was initiated by the journalist (presumably, writing) is not expressed by any of the sentence's constituent expressions. Rather, the implicit meaning is the result of an inference – an operation which is not syntactic in nature – to the effect that, because journalists typically write articles, (2) most likely means that the journalist began writing (or typing et sim.) the article. Issues relating to semantic ambiguity will be briefly touched upon below. Sentences like (2) are usually referred to as instances of 'complement coercion'. Their interpretation is thought to involve some form of 'enriched composition', resulting in a model of the eventuality that includes also a representation of the activity [**171, 98**].

1.1.1. *Reading-time and eye-tracking research into complement coercion.* McElree et al. [**137**] were the first to report longer self-paced reading times at the noun (e.g. 'article') and at the noun + 1 (e.g. 'before') positions in coercing sentences (2) compared to controls (1). McElree et al. [**136**] report that coercing constructions are interpreted less accurately, that is, more errors were made during an on-line sensicality judgment task, and more slowly than control sentences. Traxler et al. [**205**] observed that coercions result in longer total reading times as well as eye fixations on and regressions from the noun phrase region compared to control sentences. To rule out the possibility that processing differences are due to some property of verbs like 'begin', for instance their syntactic polyvalence characterized above, or semantic ambiguity (on this point, see [**204**] discussed below and [**136, 62**]), the authors compared (1) and (2) to sentences like (3), in which an event-denoting NP occurs as a complement:

(3) The journalist began the meeting before his coffee break.

Here, the activity initiated by the journalist is expressed by the noun phrase 'the meeting'. If processing difficulty is due to the verb, then there should be no difference in reading times between (2) and (3). If however increased processing costs are due to recovering or inferring silent meaning, (2) should still result in longer reading times compared to (3). Complements denoting entities ('article') are indeed associated with longer second pass and total reading times in the critical NP region compared to event-denoting complements ('meeting'). This suggests that readers experience difficulty only when the full event sense is not encoded in the lexical items, but must be inferred.

Traxler et al. [**204**] propose that coercion involves four processing steps which unfold sequentially after the reader encounters the complement. Following type-shifting accounts of coercion, such as Pustejovsky's [**171**], they argue that the complement 'the article' violates the verb's semantic requirements, as 'begin' should be followed by an expression denoting an event such as 'the meeting' or 'writing the article'. Hence, enriched composition consists in shifting the semantic type of the complement NP such that an entity-denoting expression is used to refer to an event instead. Here are the proposed processing steps:

1. The noun's lexical entry is accessed and an attempt is made to integrate the stored senses into the semantic representation on-line;
2. The mismatch between the verb's semantic requirements and the stored lexical senses of the noun trigger a coercion computation (Steps 3 and 4);
3. Properties associated with the noun and properties associated with the discourse are used to infer the correct event sense related to the entity-denoting complement;
4. The event sense is incorporated into the semantic representation of the discourse by reconfiguring the semantic representation of the complement from that of the entity to that of the related event.

The authors argue that the processing difficulties observed in earlier experiments [**137, 205, 136**] result from Step 4 of the proposed model. In order to support this claim, they designed and tested two types of stimuli. The first type (Type 1) were such that a context anticipated the event sense of the coercing noun phrase in the target sentence:

(4) a. The soldier was drinking in a bar next to the base. (Context)
b. In the afternoon, he began a bottle of imported scotch. (Target)

In target sentences, Step 3 should be relatively quick as the comprehender can immediately home in on the event sense provided by the context ('drinking'), and need not spend time sorting through the variety of associated event senses. Therefore, if Step 3 is the costly step, processing time should be reduced or eliminated with Type 1 sentences. In the second type of stimuli (Type 2), context and target sentences contained an identical coercing expression ('the novel'), therefore the context was expected to prime the coercion process (shifting the semantic type of the noun phrase) in the target sentence:

(5) a. The schoolboy began the novel about the mafia family. (Context)
b. Once he began the novel he forgot all about doing his homework. (Target)

If coercion costs are eliminated for Type 2 and not for Type 1 sentences, Traxler et al. reason, this would indicate that Step 4 is the costly step in the coercion process.

The results show that there are more eye fixations and regressions in the noun phrase and post noun-phrase regions in target sentences of Type 1 stimuli than in the same regions in control sentences. Traxler et al. conclude that Step 3 is not the costly step since processing difficulty was not significantly reduced when the context provided the appropriate event sense. The results from Type 2 sentences do not show any significant difference in fixations and regressions in the noun and post-noun phrase regions compared to control sentences. Based on this finding, the authors argue that Step 4 is the costly step.

1.1.2. *An MEG study of complement coercion.* The complexity of coercing constructions was recently demonstrated with MEG. Pylkkänen and McElree [**173**] recorded event-related fields (ERFs) while participants read sentences like (1), (2) and (6) word-by-word.

(6) The journalist astonished the article before his coffee break.

Multiple-source models were used to account for the event-related fields elicited by critical nouns. Semantically anomalous nouns modulated activity in left temporal cortex, resulting in larger source amplitudes compared to coercing and neutral nouns that peaked around 350 ms following noun onset. In the ERP literature, after Kutas and Hillyard [**117**], semantic anomalies are known to increase the amplitude of the N400, a negative component elicited by every open class word, which appears around 250 ms, peaks at about 400 ms, and lasts until approximately 550 ms following the onset of the word. More recent work [**186, 88, 82**] has suggested that the left superior temporal cortex is one of the main sources of the N400. Therefore, the M350 and the N400 might share neuronal generators in left temporal cortex. By contrast, the ERFs elicited by coercing nouns seem to be best accounted for using sources in ventromedial prefrontal cortex [**173**]. This anterior midline field (AMF) is more sensitive to coercion, and less to semantic anomaly, and vice versa for the left temporal field.

1.1.3. *The present ERP study.* The studies just reviewed show that coercing expressions give rise to some form of on-line enriched composition. However, connecting these findings to current knowledge of the processing mechanisms (as opposed to the functional anatomy) of language in the brain is problematic. That is because a sizable portion of that knowledge comes from ERP studies, whereas there currently is no EEG work on complement coercion. Linking the MEG findings of Pylkkänen and McElree [**173**] to the ERP literature on semantic processing is quite possible, but not straightforward.

It is becoming increasingly clear that the N400 [**117**] is not the whole story as far as the neurophysiological correlates of semantic integration are concerned. Still, the N400 remains a reliable measure of semantic processing load and an important term of comparison when brain correlates of different aspects of semantic composition are sought. Thus, experiments investigating enriched composition would benefit from being able to directly compare the responses evoked by, say, complement coercion with a standard N400. Such a comparison is possible on the basis of MEG data only indirectly. In particular, source reconstructions of ERF data may be consistent with several patterns of ERP effects. Coercing and anomalous nouns modulate responses at different MEG sources, and this arguably implies functional differences between processing semantic anomalies and complement coercions. However, AMF modulations were reported only very recently [**173**], they do not seem to surface with equal strength across subjects and experimental paradigms [**173, 11**], and they are not found in all putative cases of enriched composition, such as concealed questions [**85**]. ERPs afford a direct comparison with the well-established N400, hence they may provide more reliable evidence for/against a functional dissociation between processing semantic anomalies and complement coercions.

2. Method

2.1. Participants. Twenty-one right-handed native speakers of English (9 women, age range 18-34) participated in the ERP study. Varieties of English spoken by subjects included American, Canadian, Australian and British dialects. Participants had normal or corrected-to-normal vision, and no history of neurological or behavioral disorders. Subjects were paid for taking part in the study. Three subjects (1 woman) were left out of the final analysis due to a high number (> 20%) of trials contaminated by artifacts.

2.2. Materials. The set of materials consisted of 159 triplets like the following:

Coercing: The journalist *began* the article before his coffee break.
Anomalous: The journalist *astonished* the article before his coffee break.
Neutral: The journalist *wrote* the article before his coffee break.

The coercing verbs used and their frequencies within the stimulus set were: *begin* (15), *finish* (27), *start* (25), *try* (19), *attempt* (3), *master* (12), *endure* (6), *complete* (20), *enjoy* (28), *manage* (1) and *resist* (3) (see [**137, 205, 204**] for similar choices). From the original set of 159 triplets, 3 test versions were constructed, each containing 53 items per condition.

Each version included also 150 fillers, identical across versions. Fillers were sentences of varying structure, content and length, and included coercing and anomalous sentences with the critical VP in different positions in the sentence structure. Coercing and anomalous fillers were added so that readers could not anticipate where in the sentence they would encounter the coercing or anomalous noun. Each participant read 159 critical sentences (53 per condition) and 150 fillers.

The default event sense associated with each coercing VP was determined using a fill-in-the-blank questionnaire administered to 25 subjects, with 160 items like: 'The journalist began ——— the article before his coffee break'. Conflating synonyms and near-synonyms, on average 75% of the respondents selected the same event sense, which was then used as the verb in the corresponding neutral items.

A second pre-test was aimed at matching verbs' mean frequencies across conditions. Using the CELEX English corpus [**4**], lemma frequencies per million words were as follows: coercing, 314.7; anomalous, 255; neutral, 299.8 (all comparisons using T-tests, $P > 0.05$). Word-form frequencies per million words were: coercing, 31.1; anomalous, 26.6; neutral, 23.6 (all comparisons using T-tests, $P > 0.05$). Verb length was matched across conditions: coercing, 6.74; anomalous, 6.52; neutral, 6.44 (all comparisons using T-tests, $P > 0.05$).

A third pre-test was carried out to match the cloze probabilities of the object nouns in neutral and coercing sentences. The cloze probability of a particular word is defined as the frequency with which it occurs in the overall response set at a particular position in a sentence. A fill-in-the-blank test was administered to 40 participants, using 159 sentences per condition: 'The journalist began the ———' or 'The journalist wrote the ———'. Nouns were selected for use in the ERP stimuli that were equiprobable in the neutral and coercing cases based on the survey results. The mean cloze probabilities after adjustment were 0.061 for the coercing and 0.086 for the neutral condition (T-test, $P > 0.05$). Cloze probabilities were 0 for the anomalous case: nouns (such as 'article' following 'The journalist astonished the') were strongly dispreferred and therefore likely to be processed as anomalous.

2.3. Procedure. After applying the electrodes (see 2.4), participants were conducted into the experimental booth and were asked to sit in front of the video monitor. Subjects were given written instructions asking them to avoid eye blinks and movements during the presentation of the stimuli. They were also directed to read the sentences carefully and were informed that the experimenter would ask them some questions at the end of the session. Stimulus sentences were presented one word at a time for 300 ms per word, followed by 300 ms of blank screen before the onset of the next word. Words were presented in white on a black background in the center of the screen. After 1000 ms of blank screen following the offset of the last word of each sentence, an asterisk mark appeared for 1500 ms, during which subjects were allowed to blink.

2.4. Recording. EEG and EOG signals were recorded using Ag/AgCl electrodes. The EOG was measured from 4 electrodes: one at the outer canthus of each eye, one below and one above (Fp1) the left eye. The EEG was measured from 28 electrodes, arranged according to American Electrophysiological Society conventions: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2. Two additional electrodes were placed on the left and right mastoids, the former serving as the reference during the measurement. All EEG and EOG electrodes were re-referenced off-line to a linked mastoid. EEG electrodes were attached to an elastic cap, whereas EOG and reference electrodes were applied using two-sided adhesive decals external to the cap. Electrode impedance was kept below 5 kΩ throughout the experiment. The EEG/EOG was amplified by a multichannel BrainAmp DC system, with a 500 Hz sampling rate, a low pass filter set at 100 Hz and a 10 s time constant.

2.5. Data analysis. EEG data were analyzed using the MATLAB package FieldTrip.[1] The following transforms were applied to each subject's dataset. Segments corresponding to the determiner and the noun were extracted from the raw EEG with an interval of 200 ms preceding and 1000 ms following stimulus onset. Baseline correction used the 200 ms interval preceding the onset of the stimulus. Artifact rejection was based on two FieldTrip functions: the first detects and rejects all trials that contain activity exceeding a threshold of $\pm 100\ \mu V$; the second identifies and discards trials contaminated with eye movements or blinks by means of thresholding the z-transformed value of the raw data in the EOG channels, preprocessed using a band-pass filter of 1-15 Hz.

For the analysis of ERPs, a 30 Hz low-pass digital filter was applied to the segmented, artifact-free data. ERPs were obtained for each participant by averaging over trials in each experimental condition. A 5 Hz low-pass filter was used to produce the waveforms shown in figure 1. Topographical plots and statistical analyses of ERP data are however based on the 30 Hz low-pass filtered data.

To test for significant differences between conditions in the time windows of interest, we performed repeated measures ANOVA with two within-subject factors: Experimental Condition (Coercing/Control/Anomalous) and Electrode Site (28 levels, one for each scalp electrode). For the contrasts between condition pairs, a similar analysis was conducted, in which the factor Experimental Condition had two levels only. ANOVAs were carried out in two time windows: 300-500 ms, corresponding to the latency of the N400; and 700-1000 ms, based upon a visual inspection of ERP waveforms (see below).

To characterize the temporal profiles of the ERP effects elicited by coercing, anomalous and neutral nouns, we used a non-parametric randomization procedure [**129, 130**] which uses mean amplitude (μV) values in each condition for time bins of 100 ms, starting from the onset of the critical word and ending 1000 ms after. The outcome is a number (possibly 0) of clusters of electrodes in which differences between two conditions are significant in each time bin, the sum of T-statistics in that cluster, and Monte Carlo estimates of P-values.

3. Results

An inspection of ERP waveforms elicited by the determiner (data not shown) and the noun (figure 1) revealed no differences between conditions in the early ERP components such as the N1 and P2 (in both cases: 0-150 ms, $P > 0.1$; 150-300 ms, $P > 0.1$). The ERP modulations evoked by the determiner do not differ across conditions in later time windows either. This conclusion is supported by the results of ANOVA statistics on the standard N400 and P600 time windows (data not shown in figure; 300-550 ms, $P > 0.1$; 550-900 ms, $P > 0.1$).

[1] For more information, see `http://fieldtrip.fcdonders.nl/`

Differences emerge only around 300 ms following the onset of the critical noun (figure 1). The noun in the anomalous condition elicits a larger N400 compared to the same lexical item appearing in the control condition. Similarly, the coercing noun evokes what appears to be a larger N400 (but see below) compared to the control noun. Both effects start around 300 ms following word onset, peak approximately 100 ms later, and have a centro-parietal distribution as is characteristic of the N400. In both cases, the effect is statistically reliable in the standard N400 time window: 300-550 ms (table 1). In this interval, there is no significant difference between the effects evoked by anomalous and coercing nouns.

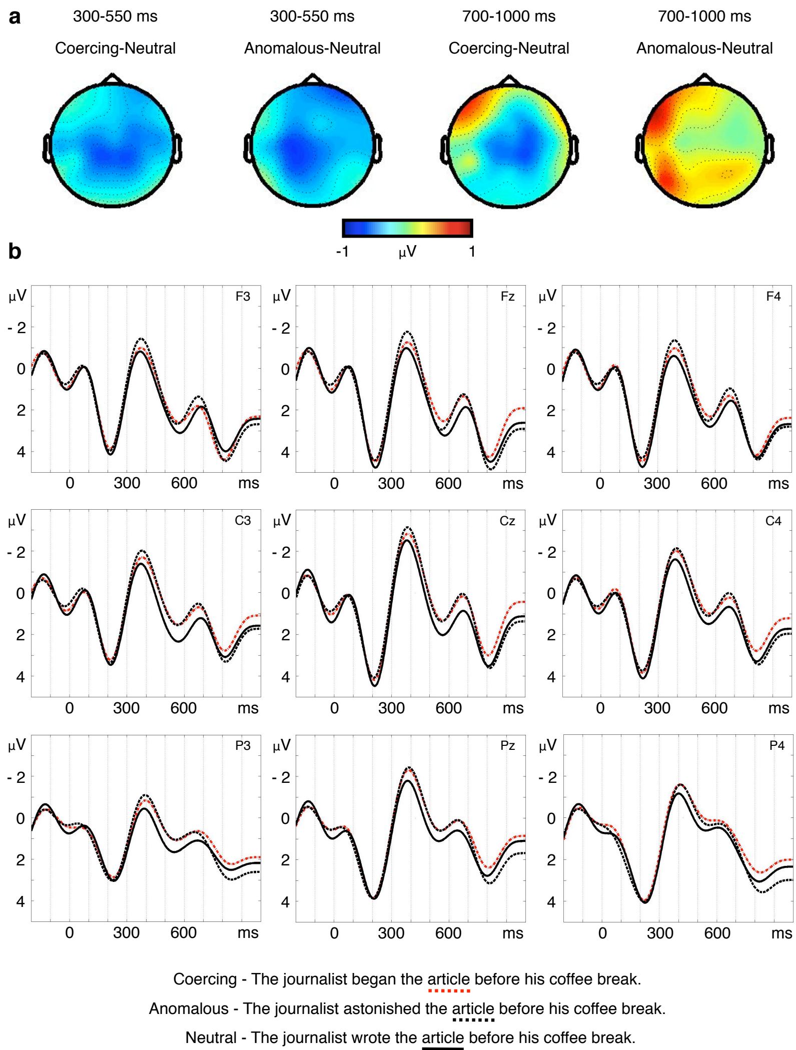


FIGURE 1. (a) Grand-average (N=18) topographies displaying the mean amplitude differences between the ERPs evoked by coercing *vs.* neutral nouns and anomalous *vs.* neutral nouns, in the 300-550 ms and 700-1000 ms regions following noun onset. (b) Grand-average (N=18) waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the noun. Negative values are plotted upward.

An examination of ERP waveforms (figure 1-b) suggests that brain potentials specific to coercion, in which the coercing condition differs from both the neutral and the anomalous cases, surface in a later time window, that starts at about 700 ms from noun onset. Coercing nouns evoke a negative-going shift with a central scalp distribution (figure 1), lasting until the end of the epoch (1000 ms). The topography of this negativity resembles that of the ERP effect evoked by coercing nouns in the 300-550 ms interval (figure 1-a, compare the first and third plots). This effect is statistically highly significant compared to the control condition (table 1), and approaches significance compared to the anomalous case ($P < 0.1$, see below for a more sensitive test using cluster-and-randomization analysis). The difference between anomalous and control nouns is, in this time interval (700-1000 ms), not significant.

	300-550 ms	700-1000 ms
Condition	$F(1,17)=3.948$ *	$F(1,17)=13.600$ ***
Coercing *vs.* Neutral	$F(1,17)=4.6245$ *	$F(1,17)=14.925$ ***
Anomalous *vs.* Neutral	$F(1,17)=4.5411$ *	$F(1,17)=2.8552$
Coercing *vs.* Anomalous	$F(1,17)=0.0441$	$F(1,17)=3.7095$
Significance codes: $p<0.001$ '***'; $p<0.01$ '**'; $p<0.05$ '*'		

TABLE 1. Summary of ANOVA statistics for the ERP analysis.

Due to the resemblance in scalp topography between the earlier and later negativities evoked by coercion (figure 1-a), as well as to the seemingly continuous negative shift in the waveforms (figure 1-b), it may be suggested that the earlier ERP shift persists as a *sustained negativity* until the end of the epoch, or that the N400 effects evoked by downstream words are also larger. Cluster-and-randomization analyses indicate that this is most probably the case: anomalous nouns, as compared to neutral nouns, give rise to statistically significant activity over a centro-parietal cluster of electrodes (sites: C3, Cz, CP5, CP1, P3, Pz) limited to the 300-500 ms interval (see table 2); coercing nouns modulate activity in similar scalp regions (electrodes: C3, Cz, C4, CP1, CP2, P3, P4), starting at 300 ms and lasting until 1000 ms instead (table 2). The difference between anomalous and coercing nouns was significant after 700 ms (table 2), providing a more sensitive statistical test than ANOVAs [**129, 130**] (compare the Coercion *vs.* Anomalous contrast in tables 1 and 2).

4. Discussion

Our results support the claim that complement coercion gives rise to definite processing costs – a claim that is now backed by reading-time, eye tracking, MEG and EEG evidence. It can therefore be concluded that, with respect to neutral items, coercing sentences require additional computations (whatever their exact nature), as is shown by reliable differences in the Coercion *vs.* Neutral contrasts. Furthermore, the insertion of an additional semantic element (corresponding to the silent activity) in coercing constructions leads to downstream unification costs, as is shown by the Coercion *vs.* Anomalous contrast in later regions. The

ERP correlates of coercion differ from those of semantic anomaly in their temporal profile at least. None the less, it cannot be concluded that there is a *qualitative* difference between the sustained negativity evoked by coercion and the N400 elicited by anomaly, because the former could also be taken as a sequence of N400 effects evoked by the coercing noun and more downstream words. But it does seem that this negative shift should be attributed to coercion, because it can hardly be explained by other factors, like the repetition of coercing verbs across stimulus sentences. It was shown by Rugg [**182**] that repeated words evoke an early (200 ms) transient negativity and a later, topographically widespread and temporally sustained positive shift. Consistently with these findings, Bresson et al. [**12**] reported that repetition reduced the amplitude and shortened the duration of the N400 component. Our study revealed rather the opposite pattern of effects.

	Coercing *vs.* Neutral	Anomalous *vs.* Neutral	Coercing *vs.* Anomalous
300-400 ms	$T(16)$=-54.874 **	$T(16)$=-37.412 *	
400-500 ms	$T(16)$=-61.287 **	$T(16)$=-34.018 *	
500-600 ms	$T(16)$=-52.658 **		
600-700 ms	$T(16)$=-50.151 *		
700-800 ms	$T(16)$=-44.456 *		$T(16)$=-20.135 *
800-900 ms	$T(16)$=-53.645 **		$T(16)$=-35.419 **
900-1000 ms	$T(16)$=-64.083 **		$T(16)$=-63.260 **

Significance codes: p<0.01 '**'; p<0.05 '*'

TABLE 2. Summary of cluster-and-randomization statistics for the ERP analysis. The first significant effects occurred at 300-400 ms. Empty cells denote the absence of significant clusters.

4.1. Simple and enriched composition. In this chapter we have presented ERP data that, in line with earlier reading-time, eye-tracking and MEG studies, suggest that enriched semantic composition is involved in processing sentences in which verbs like begin are used transitively with two NP arguments, like in 'The journalist began the article'. It therefore seems that the neural reality of complement coercion, which has now been established with MEG [**173**] and EEG, challenges linguistic theories that analyze these constructions in a strictly compositional manner [**172**] (see also chapter 3).

Theories exist that model the coercion process in a non-compositional manner, where world knowledge is involved alongside syntax and lexical semantics. One such proposal is based on *unification*, a mechanism adopted by several linguistic accounts of syntactic and semantic structure assembly [**106, 221, 212**]. Equally viable approaches have been developed that do not rely explicitly on unification [**171, 98**], but what makes that concept appealing here is that accounts of language processing inspired by neurobiology are trying to bridge the gap between linguistics and neuroscience by means of the notion of unification [**71, 72**]. Complement coercion can be described in a unification-based theory of tense, aspect and event structure, and that is the Event Calculus [**212**] applied in earlier chapters. The reader may recall that the formalism is type-free, and is therefore alternative to analyses based on types and type-shifting [**171**], which earlier experimental research on complement coercion made reference to [**137, 205, 204, 136**].

In the Event Calculus, the verb 'begin' and its synonyms are represented by the clause *Initiates*(*start*,a,t), which means that a starting event (*start*) initiates the activity a at time t. The variable a has to be unified with material provided by the context or world knowledge, that is, a VP ($a = write$) or a nominalized event ($a = meeting$). This formalism attractive also due its cognitive plausibility, discussed in greater detail in [**212**]. Semantic representations for VPs and the unification operation find their motivation in an hypothesized evolutionary connection between the semantics of tense, aspect and event structure on the one hand, and planning on the other [**191**] (chapter 1). The Event Calculus is in fact a planning formalism, which has been applied in the design of intelligent agents in robotics, and to the semantics of temporal expressions in linguistics [**212, 83**]. Three arguments justify our preference for unification over type-shifting as a framework in which to try to make sense of our ERP data: a connection with the notions of unification and binding in neuroscience [**71, 72**], cognitive plausibility through a link with planning and an ontology that does not involve types, and success in modeling linguistic phenomena.

It should be emphasized that the ERP data reported here do not favor unification over alternative accounts of coercion such as Putesjovsky's [**171**] type-shifting notion, nor was the present ERP study designed to verify predictions suggested by the unification model. However, and conversely, the unification framework can be used to shed some light on our ERP data. The Event Calculus, in particular, suggests that coercing sentences require more complex unification operations than neutral sentences, as an extra inference step has to be made in order to unify the variable a with a plausible activity. The goal of this (abductive) inference step is precisely to find out an activity that fits semantically the given discourse context. As can be seen, unification formalizes the traditional notion of integration of lexical items into a sentence or a discourse model. Therefore, to the extent that the N400 effect can be regarded as a signature of semantic integration (see below), the N400-like shift elicited by the coercing noun may be associated with the more complex, inference-driven unification leading to a richer event structure. The sustained negativity, or the sequence of N400 effects involving also more downstream words, could reflect difficulty in integrating lexical items following the noun into a more complex semantic representation of the unfolding sentence. The type-shifting account formalizes the traditional notion of integration in different terms, but in this case it may not lead to an entirely different account of the data.

4.2. Enriched composition and lexico-semantic anomaly. Pylkkänen and McElree correctly remark that the nature of the N400 as an index of integration difficulty has so far not been demonstrated [**173**]. Experimental evidence points in the direction of two competing accounts: the first, which postulates that the N400 is directly correlated with the difficulty in integrating the meaning of the upcoming word in the sentence model [**208, 209, 79, 155, 52, 81, 148, 127**]; the second, which proposes the N400 is inversely correlated with the amount of pre-activation a word receives from the preceding context [**117, 118, 216, 217, 93, 76, 47, 115, 35**]. It is important to avoid setting up a false opposition, because these two accounts merely highlight different processes that may underlie the generation of the N400, but are in no way alternative and mutually exclusive theories. In fact, a unified view that fits most data seems conceivable, in which different roles are assigned to each cerebral hemisphere: predictive semantic coding is a left hemisphere mechanism, whereas the right hemisphere contribution is strictly postlexical in nature, thus contributing to the integration of word meanings, that is, to unification [**47, 115, 75**].

Within this unified framework for the N400, our ERP data seem to favor an explanation based on integration. In particular, if the N400-like negativity elicited by the coercing noun were to be accounted for in terms of lexical pre-activation, then it seems the unification story sketched above (or any other interpretation based on integration, for that matter) should be abandoned. According to the pre-activation view, the amplitude of the N400 component is inversely correlated with the amount of lexico-semantic priming a word receives from the preceding sentence or discourse context [**47, 115**]. One driving factor behind pre-activation is semantic relatedness, so 'article' should be pre-activated more strongly by the fragment 'The journalist wrote the' as compared to 'The journalist astonished the' and 'The journalist began the'. However, the cloze probability data reported in section 2.2 indicate that there is very little difference between neutral ('the journalist wrote the') and coercing ('the journalist began the') contexts as far as the expectancy of the noun 'article' is concerned. This seems to challenge an account of the ERP correlates of coercion that is based on lexical pre-activation alone. As a result, it lends some plausibility to the integration/unification view.

As we noted above, complement coercion requires inferring a plausible event sense based on world knowledge, and integrating that (via unification or type-shifting) into the event structure of the VP began the article. This process appears to be non-compositional, and has two brain correlates: (a) a modulation at sources different from those that generate the N400, as the AMF data by Pylkkänen and McElree [**173**] show; (b) a modulation of the N400 at the coercing noun and more downstream regions, leading to a sustained negative shift. Our experiment does not license the conclusion that the neural processes underlying enriched composition are qualitatively distinct from those supporting standard semantic composition and the processing of semantic anomalies. We did find a specific ERP effect of coercion, but this differed from a standard N400 only in duration: a sustained negativity, or a sequence of N400s to downstream words was found, with a similar scalp distribution as a standard N400 effect. Therefore, contrary to what one might have expected on the basis of the MEG study, the ERP data did not distinguish between anomaly and coercion in the typical N400 window, which is the time interval in which the AMF was found (350-500 ms). Based on the data available so far, it is impossible to say how exactly, if at all, the AMF, the N400 and the sustained negativity evoked by coercion could be related. Further MEG, EEG and possibly fMRI studies into enriched composition seem necessary to address this issue. Additional experimental testing might also be useful to address the conjecture, proposed by Pylkkänen and McElree [**173**], that the physiologic responses specific to coercion (the AMF in their report) may be generated in medial prefrontal cortices, which have been associated with perspective taking and inferring others' mental states. The interpretation of coercing sentences, and the abductive inference to find out a plausible activity, might involve taking the perspective of the agent depicted in the sentence (that is, the journalist). EEG data might reveal whether the N400 is sensitive to varying degrees of complexity in perspective taking during sentence processing, and fMRI data might give us a more precise picture of the brain structures recruited by non-compositional meaning assembly.

To conclude, together with the ERP results reported in chapter 5, these data provide some evidence that the brain computes and updates models of the input in an incremental, yet *non-monotonic* and *non-compositional* manner. Negative shifts in the ERPs, N400s as well as slower potentials, constitute valuable on-line measures of such processes.

Conclusions

As stated in the foreword, the aim of this book was to combine theoretical and experimental investigations on meaning processing at the sentence and discourse levels. It is now time to stand back, consider how the resulting picture looks like, and what are its implications for research in linguistics and neuroscience. In this final section we will first provide a synthesis of current knowledge of semantic computation in the human brain, with some emphasis on the hypotheses and results reported in previous chapters. Next we will introduce problems to be addressed in future work, and take a few steps towards a neurobiology of meaning.

One source of difficulty in providing a systematic account of semantic computation is due to immediacy and its corollary incrementality. These two principles are well-established in psycholinguistic research, but they are notoriously hard to formalize. The notion that the system handles first the information that becomes available first implies that each utterance, and not just each sentence, is processed differently. This makes any attempt of generalizing over the processing steps involved in computing complex meanings particularly hard. The temporal and logical dependencies of unifications in phonology, morpho-syntax, semantics and pragmatics are indeed utterance-dependent. For instance, the different positions of the temporal adverbial in

(1) a. Last sunday Vincent painted the window frames of his country house.
    b. Vincent painted the window frames of his country house last sunday.

are likely to have an effect on how temporal constraints are set up and solved, and therefore may lead to different ERP responses. However, ERP components define temporal windows in which processes of the same type occur routinely. It is on these regularities that theories of language processing should focus. This would allow us to address three basic questions: (a) What are the goals of computation in semantic structure assembly? (b) What is retrieved from semantic memory given the input? (c) How does computation take place? The reader may recognize (a) and (b) as Marr level-1 (computational) issues, and (c) as an instance of a level-2 (algorithmic) issue.

Our answer to (a) is that brain systems compute, and incrementally update, a *minimal model* of the input. The definition of minimality adopted here relies heavily upon the work by Stenning, van Lambalgen and Hamm [**212, 196**], and is based on closed world reasoning in logic programming (negation as failure) and program completion as a means of deriving minimal models. Other notions of minimality are available, for example in the framework of Optimality Theory and Harmonic Grammar [**7, 188**]. If these different definitions lead to different processing predictions, it makes sense to treat them as testable hypotheses on the nature of semantic representations in the brain. Additional, and possibly crucial constraints may be suggested by knowledge of the biophysical implementation of discourse models as equilibrium states of the cortical networks involved in computing complex meanings [**72**]. Regardless of the details concerning the proper definition of minimality, it seems necessary to postulate that the brain resorts to some minimization procedure for obtaining discourse models: language understanding is not based on *all* models of what is said, for if that were the case, human working memory capacity would soon be exceeded for even the simplest of utterances. ERP data that speak to this problem, however indirectly, have been presented in chapter 5. Further research is needed to understand what minimization mechanisms the brain actually employs, what kind of information is left out of minimal models, and why.

Our answer to (b) is somewhat less dependent on ERP results. The reader may wonder why our discussion of semantic processing in the brain does not start with hypotheses on the nature of elementary representations stored in memory. Here a different approach was chosen. Processing predictions were derived from formal semantic theories, and these are typically more focused on sentences and discourses than on the lexicon. That explains why we had not much to say about lexical processes. In a way, however, the choice of unification as a basic combinatory mechanism does suggest a particular view of stored representations as constraint-based data structures. This notion is explicit in chapter 4. Just like the binding of foot and root nodes in Vosse and Kempen **[221]**, the unification of temporal variables that takes place during tense processing is best treated as a constraint-solving task. Constraints are essential for computing with languages that involve variables and individual constants. However, there is more to atomic semantic data structures than constraints in the formal sense, and that is conceptual content. Lexical meanings must include *constraints*, that code for dependencies between denotations of different parts of speech (individuals, times etc.), and *predicates*, representing the features that jointly constitute the conceptual content of an expression. The language of the Event Calculus makes this distinction too.

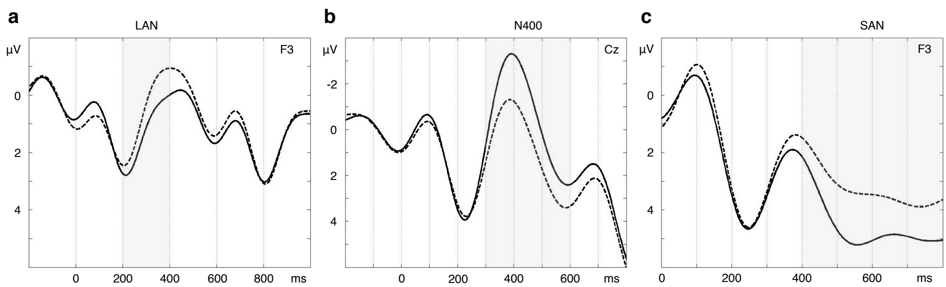


FIGURE 2. Grand-average waveforms displaying ERP effects correlated with semantic complexity. (a) Left Anterior Negativity: dotted line, verb tense violation; continuous line, control (chapter 4). (b) N400: continuous line, lexico-semantically unrelated word; dotted line, control (chapter 5). (c) Sustained Anterior Negativity: dotted line, semantic recomputation; continuous line, control (chapter 5).

Let us now address (c), and see how the proposed theory accounts for known language-related ERP effects. Within the first 200 ms following the moment at which a stimulus word impinges on the peripheral visual or auditory systems, physical changes in the environment are recognized as linguistic input. Within the following 200 ms, neural activation spreads to temporal cortices, activating phonologic, syntactic and semantic representations associated with the input. Retrieved meanings are necessarily partial structures (chapter 3), although selection mechanisms are still largely unknown. ERP data (figure 2-a) suggest that, during this time interval ($\approx$ 200-400 ms from stimulus onset) the system tries to satisfy constraints that are bound to morphological features of words, such as word category, tense, case, number, and gender. As we have argued in chapter 4, regardless of the level at which processing occurs in these cases (possibly at that of word form, rather than semantic representations), the LAN demonstrates that semantic constraints set up by the preceding material are taken into account when processing each incoming word. This shows that some form of semantic computation involving variables or individual constants (e.g. for time instants or intervals) takes place within the first 400 ms from stimulus onset.

Once morphologically-bound constraints (word category, agreement etc.) are satisfied, aspects of semantic computation involving predicates (conceptual content) become more prominent. Between 300 and 500 ms from stimulus onset, the system tries to unify semantic features across parts of speech. This can be seen as a second-order unification of properties, analogous to the first-order unification of variables or constants (regimented by constraints) that sets on around 200 ms from word onset. The N400 (figure 2-b) is an electrophysiological marker of these processes. For instance, in the context of the fragment 'The girl was writing a letter when her friend spilled coffee on the', the word 'tablecloth' evokes a larger N400 as compared to 'paper'. This might be because 'paper' is semantically quite related to 'writing a letter', so some feature in its lexical representation can be easily unified with some other feature in the representation of the VP. This cannot be done (so easily) for 'tablecloth', hence the larger N400. Consider another example: the word 'article' in the three different contexts 'The journalist wrote/began/astonished the'. Unification of higher-order properties seems easier for 'article' following the fragment 'The journalist wrote the' as compared to the other two cases, because there seems to be very little in the meaning of 'begin' and 'astonish' that is shared with 'article'. These unification attempts might involve a word and its preceding sentence or discourse context (as in the cases just considered), world knowledge, perceptual input delivered from other modalities, co-speech gestures etc. – recall the various situations considered in chapter 3. The N400 effect suggests that the brain tries to construct a semantic representation in which features are connected *across* constituent expressions. In our theory, this is achieved by repeated applications of unification. Modulations of the N400 amplitude are produced when unifications are harder to compute, or more unifications are required to arrive at a discourse model.

Toward the end of the N400 time window (around 550 ms), the current minimal model has been updated with information brought about by the relevant content word. The bulk of the computation so far has been semantic, except for the early processing stages in which morphological information is also prominent. Syntactic unification might also occur before $\approx$ 550 ms. However, the P600 shows that *disrupting* syntactic processing, if only temporarily as in garden-path constructions, has somewhat later consequences (550-900 ms). This might be taken to suggest [**201**] that syntactic unification is a relatively late process, which secures the unifications proposed by the semantics. This stands in contrast with the received picture of language processing (syntax first, semantics next), but it does fit with important ERP data such as the 'semantic attractions' [**111**] discussed in chapter 3. In approximately the same time lapse ($\approx$ 400-800 ms), another set of operations takes the unifications proposed by the semantics as input, but this time with a different purpose – evaluating the consequences of unification for the discourse model. Here inference processes come to the fore, and possible ERP signatures may be the SFN (sentence-final negativity) of chapter 4, the SAN (sustained anterior negativity) of chapter 5 (figure 2-c), and the prolonged negative shift of chapter 6. These are different effects physiologically (as is testified by different temporal profiles and scalp distributions) and, as a consequence, functionally. However, they all seem to support processes that belong to the same family, namely a late, global adjustment of the discourse model so as to fully and consistently represent the input. It remains unclear whether such processes can also take place in the N400 time window, and if so, to what extent. This relates to the fundamental issue whether the N400 is the *only* ERP index of semantic processing in the brain, and sustained negativities are rather an effect of working memory or attentional resources modulated by the task. Further research is necessary to address this issue.

The picture just presented makes a distinction between three, partly overlapping time windows in which functionally distinct processes take place. From 200 to 400 ms following word onset, attempts are made to satisfy the constraints set up by morphologically-bound features. Here semantics comes into play whenever word forms have to be evaluated based on the preceding semantic material, as is the case for tense violations. These processes can be modeled along the lines of the agreement check of Vosse and Kempen [**221**], which takes the form of feature matching in morpho-syntax (e.g. case) and the unification of individual variables occurring in semantic constraints (tense). This early processing stage is indexed in the ERPs by (E)LANs. Between 300 and 550 ms, higher-order semantic features of words are accessed, and attempts are made at connecting these across retrieved lexical structures. This process can be modeled as the unification of predicates and variables of a different sort than individuals (e.g. fluents). The N400 indexes complexity at this intermediate processing stage. From 400 to 900 ms, the outcome of first-pass semantic unification is handed over to two, seemingly parallel (given temporal overlap in the ERPs), yet interacting, components. The first validates the first-pass semantic input by unifying syntactic frames. Whenever this is not straightforward, or not at all feasible, as when there is a conflict between the proposed semantic input and the permissible syntactic unifications, a P600 is produced. The second validates the first-pass semantics by adjusting the message-level representation. If difficulty is caused by this set of operations, as when a full event sense is integrated into the model, or when the model is recomputed, a sustained negativity will be observed. Compositionality (informational encapsulation, chapter 3) is a hypothesis about the constraints holding at the interface between the lexico-semantic component and the syntactic component. In terms of ERPs, this bears on the relations between N400 and P600. Monotonicity can then be seen as a hypothesis on the computations allowed within the discourse-semantic component.

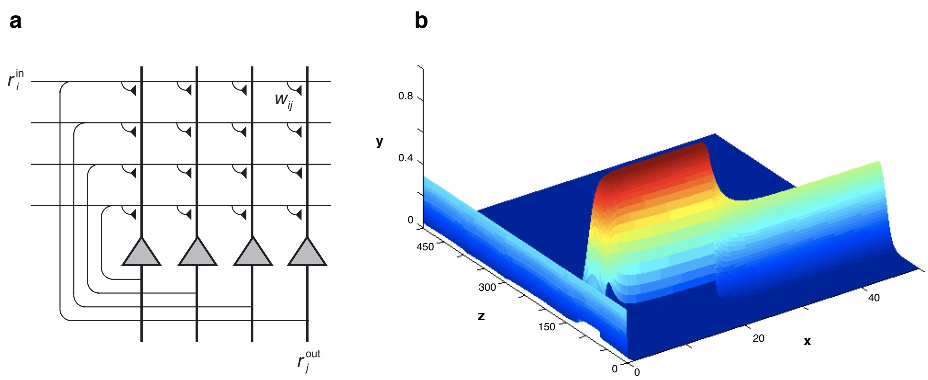


FIGURE 3. (a) Recurrent circuit from Trappenberg **[203]**. (b) Activation landscape of the recurrent network described in the main text following the presentation of a fast input spike train at $t = 0$. The x-axis represents processing-time steps. The y-axis represents activity levels in a scale-free 0-1 range. Neurons ($z = 1, ..., 500$) are represented on the z-axis.

While early-stage processes can be accounted for in a purely feed-forward architecture, a recurrent network is necessary for implementing operations at later stages. Therefore, the full computational capacity of the neocortex – which is characterized by massive recurrent connections – really comes into its own only later than ≈ 300 ms after the onset of a linguistic stimulus. This interval contains all language-related ERP components, like the N400, P600 and SAN. As a consequence, it seems any neural account of these effects should be based on an application of the recurrent architecture. Recent work using Dynamic Causal Modeling seems to point in that direction too **[63]**. Recurrent and attractor networks were mentioned on several occasions in earlier chapters. Let us now look at a simple recurrent network that can be taken as a launch pad for future research. Minimally, a network that sets out to model N400, P600 and SAN should include two sub-systems: a storage component, modeling the function of left temporal cortex; and a computation component, modeling the function of left inferior frontal cortex. This should allow us to describe the balance and the interactions between storage and computation (chapter 3) or, using Hagoort's **[72]** terminology, between memory and unification.

Our toy network is constituted by 500 leaky integrate-and-fire neurons, organized into four recurrent memory modules [**126**] (see figure 3-b for a simple 4-neuron module). These are grouped into two components: one (1 module of 100 excitatory neurons) stores learned patterns that range from intra-sentential syntactic templates to discourse structures such as logical arguments or narratives; the other (2 modules of 150 excitatory neurons each, and 1 module of 100 inhibitory interneurons) computes the stable states of the network given the input and patterns stored in the first component. Future modeling work should extend and refine a recurrent network of this kind so that it can yield theoretical curves comparable to independent components (as opposed to whole waveforms) extracted from ERPs. The key task and foremost challenge is to adjust the scale of the network's output dimensions to that of observable ERP quantities (microvolt and milliseconds), and realistically constrain their orders of magnitude. This could be done by increasing the degree of physiological realism of the network, for instance adding neurons to each module, fine-tuning time constants, and changing the network's architecture based on fMRI data and cytoarchitectonic constraints on neocortical organization. The superficial resemblance between the network's activation landscape (figure 3-b) and the SAN profile (figure 2-c) conceals the heterogeneity between biophysical ERP quantities and arbitrary scales in the network output. However, recurrent networks exhibiting fixed-point dynamics seem to meet both top-down requirements (they are suitable for approximating the semantic operators that compute minimal models [**91**]) and bottom-up requirements (they are plausible models of prefrontal cortex function [**43**]). Therefore, they are a promising platform for modeling language-related ERP effects.

In this book we have tried to lay down a path that connects semantics to neuroscience, from the philosophical and methodological considerations of chapter 1, via an examination of the processing consequences of formal semantic principles in chapters 2-3 and the ERP data of chapters 4-6, to the conjectures on recurrent networks presented above. There would be little point in arguing that this solves any of the problems associated with meaning and the brain, or that the proposed picture is accurate and complete. Nonetheless, what can be concluded is that bridging theories and processing data in the study of meaning is possible. Assuming a unified perspective is an asset, it is useful to devise an integrative methodology, sufficiently explicit to function as a guide in day-to-day work. Here methodological choices (essentially Marr's scheme, plus several add-ons specific to the task at hand) were brought to the fore since the first chapter, and applied in an almost step-by-step fashion throughout the remaining chapters. Cognitive science has been characterized by intense philosophical and conceptual debates since the very beginning, and current imaging-aided neuroscience has inherited some of that attitude. The studies presented here build upon this established feat of the behavioral and brain sciences, and contribute to keeping it alive and prominent in times when the proliferation of theories and experimental results may give the impression that there is little need for questioning the foundations of and divisions within the field.

Bibliography

[1] M. Allen, W. Badecker, and L. Osterhout, *Morphological analysis in sentence processing: An ERP study*, Language and Cognitive Processes **18** (2003), 405–430.
[2] E.J. Andrade-Lotero, *Meaning and form in Event Calculus*, Master's thesis, University of Amsterdam, 2006.
[3] L. Åqvist, F. Guenthner, and C. Rohrer, *Definability in ITL of some subordinate temporal conjunctions in English*, Studies in Formal Semantics (F. Guenthner and C. Rohrer, eds.), North-Holland, 1978, pp. 201–221.
[4] R.H. Baayen, R. Piepenbrock, and L. Gulikers, *CELEX2*, Linguistic Data Consortium, Philadelphia, 1996.
[5] S.B. Barton and A.J. Sanford, *A case study of anomaly detection: Shallow semantic processing and cohesion establishment*, Memory and Cognition **21** (1993), 477–487.
[6] R.A. Berman and D.I. Slobin (eds.), *Relating events in narrative: A crosslinguistic developmental study*, Lawrence Erlbaum Associates, 1994.
[7] R. Blutner, *Some aspects of optimality in natural language interpretation*, Journal of Semantics **17** (2000), 189–216.
[8] D. Bolinger, *Judgments of grammaticality*, Lingua **21** (1968), 34–40.
[9] G. Boole, *An investigation of the laws of thought*, Dover, 1958.
[10] E. Borg, *Minimal semantics*, Oxford University Press, 2004.
[11] J. Brennan and L. Pylkkänen, *Processing events: Behavioral and neuromagnetic correlates of aspectual coercion*, Brain and Language **106** (2008), 132–143.
[12] M. Bresson, M. Kutas, and C. Van Petten, *An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences*, Journal of Cognitive Neuroscience **4** (1992), 132–139.
[13] C. Brown and P. Hagoort, *The cognitive neuroscience of language: Challenges and future directions*, The Neurocognition of Language (C. Brown and P. Hagoort, eds.), Oxford University Press, 1999, pp. 3–12.
[14] M. Bunge, *Philosophical problems in linguistics*, Erkenntnis **21** (1984), 107–173.
[15] T. Burge, *Truth, thought, reason. Essays on Frege*, Clarendon Press, 2005.
[16] M.N. Carminati, L. Frazier, and K. Rayner, *Bound variables and c-command*, Journal of Semantics **19** (2000), 1–34.
[17] M. Carreiras, A. Garnham, J. Oakhill, and K. Cain, *The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish*, The Quarterly Journal of Experimental Psychology **49A** (1996), 639–663.
[18] P. Carruthers, *Language, thought and consciousness*, Cambridge University Press, 1996.
[19] ———, *The cognitive functions of language*, Behavioral and Brain Sciences **25** (2002), 657–725.
[20] K. Chauncey, A. Özyürek, P. Hagoort, and S. Kita, *Recognition of iconic gestures: A first gating study*, Unpublished manuscript, 2004.
[21] N. Chomsky, *The logical structure of linguistic theory*, MIT Press, 1955.
[22] ———, *Syntactic structures*, De Gruyter-Mouton, 1957.
[23] N. Chomsky, *Aspects of the theory of syntax*, MIT Press, 1965.
[24] N. Chomsky, A. Belletti, and L. Rizzi, *On nature and language*, Cambridge University Press, 2002.
[25] K. Christianson, A. Hollingworth, J. Halliwell, and F. Ferreira, *Thematic roles assigned along the garden path linger*, Cognitive Psychology **42** (2001), 368–407.
[26] B. Comrie, *Aspect*, Cambridge University Press, 1976.
[27] ———, *Tense*, Cambridge University Press, 1985.
[28] A.E. Cook, J.E. Limber, and E.J. O'Brien, *Situation-based context and the availability of predictive inference*, Journal of Memory and Language **44** (2001), 220–234.
[29] M. Cresswell, *Interval semantics and logical words*, On the logical aspects of tense and aspect (C. Rohrer, ed.), Narr, 1977, pp. 7–29.
[30] W. Croft, *The structure of events and the structure of language*, The new psychology of language: Cognitive and functional approaches to language structure (M. Tomasello, ed.), Lawrence Erlbaum Associates, 1998, pp. 67–92.
[31] W. Croft and D.A. Cruse, *Cognitive linguistics*, Cambridge University Press, 2004.
[32] P.W. Culicover and R. Jackendoff, *Simpler Syntax*, Oxford University Press, 2005.
[33] ———, *The simpler syntax hypothesis*, Trends in Cognitive Sciences **10** (2006), 413–418.
[34] M. de Kamps and F. van der Velde, *Implementation of multilayer perceptron networks by populations of spiking neurons using rate coding*, Neurocomputing **44-46** (2002), 353–358.
[35] K.A. De Long, T.P. Urbach, and M. Kutas, *Probabilistic word pre-activation during language comprehension inferred from electrical brain activity*, Nature Neuroscience **8** (2005), 1117–1121.
[36] H. de Swart, *Aspect shift and coercion*, Natural Language and Linguistic Theory **16** (1998), 347–385.
[37] J. Dever, *Compositionality as methodology*, Linguistics and Philosophy **22** (1999), 311–326.
[38] M. Devitt, *Intuitions in linguistics*, British Journal of Philosophy of Science **57** (2006), 481–513.
[39] K. Doets, *From logic to logic programming*, MIT Press, 1994.
[40] E. Donchin, *Surprise! ... surprise?*, Psychophysiology **18** (1981), 493–513.
[41] R.J. Douglas and K.A.C. Martin, *Mapping the matrix: The ways of the neocortex*, Neuron **56** (2007), 226–238.

[42] D. Dowty, *Word meaning and Montague grammar*, Reidel, 1979.

[43] D. Durstewitz, J.K. Seamans, and T.J. Sejnowski, *Neurocomputational models of working memory*, Nature Neuroscience **3** (2000), 1184–1191.

[44] M. Egg, *Beginning novels and finishing hamburgers: Remarks on the semantics of 'to begin'*, Journal of Semantics **20** (2003), 163–191.

[45] T.A. Erickson and M.E. Matteson, *From words to meanings: A semantic illusion*, Journal of Verbal Learning and Verbal Behaviour **20** (1981), 540–552.

[46] S. Featherston, M. Gross, T.F. Münte, and H. Clahsen, *Brain potentials in the processing of complex sentences: An ERP study of control and raising constructions*, Journal of Psycholinguistic Research **29** (2000), 141–154.

[47] K.D. Federmeier and M. Kutas, *Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing*, Cognitive Brain Research **8** (1999), 373–392.

[48] C. Felser, H. Clahsen, and T. Münte, *Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German*, Brain and Language **87** (2003), 345–354.

[49] F. Ferreira, V. Ferraro, and K.G.D. Bailey, *Good-enough representations in language comprehension*, Current Directions in Psychological Science **11** (2002), 11–15.

[50] F. Ferreira and N.D. Patson, *The 'good enough' approach to language comprehension*, Language and Linguistics Compass **1** (2007), 71–83.

[51] F. Ferreira and M.K. Tanenhaus, *Introduction to the special issue of language-vision interactions*, Journal of Memory and Language **57** (2007), 455–459.

[52] T. Ferretti, M. Kutas, and K. McRae, *Verb aspect and the activation of event knowledge*, Journal of Experimental Psychology: Learning, Memory and Cognition **33** (2007), 182–196.

[53] C. Fiebach, M. Schlesewsky, and A. Friederici, *Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German wh-questions*, Journal of Memory and Language **47** (2002), 250–272.

[54] W.T. Fitch, M.D. Hauser, and N. Chomsky, *The evolution of the language faculty: Clarifications and implications*, Cognition **97** (2005), 179–210; discussion 211–225.

[55] J. Fodor, *The modularity of mind*, MIT Press, 1983.

[56] E. Fonteneau, U.H. Frauenfelder, and L. Rizzi, *On the contribution of ERPs to the study of language comprehension*, Bulletin Suisse de Linguistique Appliquée **68** (1998), 111–124.

[57] S.L. Frank, M. Koppen, L.G.M. Noordman, and W. Vonk, *Modeling knowledge-based inferences in story comprehension*, Cognitive Science **27** (2003), 875–910.

[58] L. Frazier, *Sentence processing: A tutorial review*, Attention and performance (M. Coltheart, ed.), Lawrence Erlbaum Associates, 1987, pp. 559–586.

[59] G. Frege, *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Koebner, 1884.

[60] ———, *On sense and reference*, Translation from the philosophical writings of Gottlob Frege (P. Geach and M. Black, eds.), Blackwell, 1980, pp. 56–78.

[61] A.D. Friederici, *Towards a neural basis of auditory sentence processing*, Trends in Cognitive Sciences **6** (2002), 78–84.

[62] S. Frisson and B. McElree, *Complement coercion is not ambiguity resolution: Evidence from eye movements*, Journal of Experimental Psychology: Learning, Memory and Cognition **31** (2008), 1–11.

[63] M.I. Garrido, J.M. Kilner, S.J. Kiebel, and K.J. Friston, *Evoked brain responses are generated by feedback loops*, Proceedings of the National Academy of Sciences **104** (2007), 20961–20966.

[64] B. Geurts and F. van der Slik, *Monotonicity and processing load*, Journal of Semantics **22** (2005), 97–117.

[65] H. Geurts, *Executive functioning profiles in ADHD and HFA*, Ph.D. thesis, Vrije Universiteit Amsterdam, 2003.

[66] A. Goldberg, *Constructions: A construction grammar approach to argument structure*, University of Chicago Press, 1995.

[67] P.M. Greenfield, *Language, tools and the brain: The ontogeny and phylogeny of hierarchically organized sequential behavior*, Behavioral and brain sciences **14** (1991), 531–595.

[68] H. P. Grice, *Studies in the ways of words*, Harvard University Press, 1989.

[69] J. Groenendijk and M. Stokhof, *Dynamic predicate logic*, Linguistics and Philosophy **14** (1991), 39–100.

[70] P. Hagoort, *The shadows of lexical meaning in patients with semantic impairments*, Handbook of Neurolinguistics (B. Stemmer and H.A. Whitaker, eds.), Academic Press, 1998, pp. 235–248.

[71] ———, *How the brain solves the binding problem for language: A neurocomputational model of syntactic processing*, Neuroimage **20** (2003), S18–S29.

[72] ———, *On Broca, brain, and binding: A new framework*, Trends in Cognitive Sciences **9** (2005), 416–423.

[73] P. Hagoort, *The binding problem for language and its consequences for the neurocognition of comprehension*, The Processing and Acquisition of Reference (N. Pearlmutter E. Gibson, ed.), MIT Press, 2006.

[74] P. Hagoort, *The Memory, Unification, and Control (MUC) model of language*, Automaticity and Control in Language Processing (A.S. Meyer, L.R. Wheeldon, and A. Krott, eds.), Psychology Press, 2006, pp. 243–270.

[75] P. Hagoort, G. Baggio, and R.M. Willems, *Semantic unification*, The New Cognitive Neurosciences (M.S. Gazzaniga, ed.), MIT Press, 2009.

[76] P. Hagoort and C. Brown, *Brain responses to lexical ambiguity resolution and parsing*, Perspectives on Sentence Processing (C. Clifton, L. Frazier, and K. Rayner, eds.), Lawrence Erlbaum Associates, 1994, pp. 45–81.

[77] P. Hagoort, C. Brown, and L. Osterhout, *The neurocognition of syntactic processing*, The Cognitive Neuroscience of Language, Oxford University Press, 2001, pp. 273–307.

[78] P. Hagoort, C.M. Brown, and J. Groothusen, *The syntactic positive shift (SPS) as an ERP measure of syntactic processing*, Language and Cognitive Processes **8** (1993), 439–483.

[79] P. Hagoort, L. Hald, M. Bastiaansen, and K.M. Petersson, *Integration of word meaning and world knowledge in language comprehension*, Science **304** (2004), 438–441.

[80] P. Hagoort and J. van Berkum, *Beyond the sentence given*, Philosophical Transactions of the Royal Society, B **362** (2007), 801–811.

[81] L. Hald, E.G. Steenbeek-Planting, and P. Hagoort, *The integration of discourse context and world knowledge in online sentence comprehension. Evidence from the N400*, Brain Research **1146** (2007), 210–218.

[82] E. Halgren, R.P. Dhond, N. Christensen, C. Van Petten, K. Marinkovic, J.D. Lewine, and A.M. Dale, *N400-like MEG responses modulated by semantic context, word frequency, and lexical class in sentences*, NeuroImage **17** (2002), 1101–1116.
[83] F. Hamm, H. Kamp, and M. van Lambalgen, *There is no opposition between formal and cognitive semantics*, Theoretical Linguistics **22** (2006), 1–40.
[84] F. Hamm and M. van Lambalgen, *Event Calculus, nominalisation and the progressive*, Linguistics and Philosophy **26** (2003), 381–458.
[85] J. Harris, L. Pylkkänen, B. McElree, and S. Frisson, *The cost of question concealment: Eye-tracking and MEG evidence*, Brain and Language **107** (2007), 44–61.
[86] C. Haslam, A.J. Wills, S.A. Haslam, J. Kay, R. Baron, and F. McNab, *Does maintenance of colour categories rely on language? Evidence to the contrary from a case of semantic dementia*, Brain and Language **103** (2007), 251–263.
[87] M.D. Hauser, N. Chomsky, and W.T. Fitch, *The faculty of language: What is it, who has it, and how did it evolve?*, Science **298** (2002), 1569–1579.
[88] P. Helenius, R. Salmelin, E. Service, and J.F. Connolly, *Distinct time courses of word and context comprehension in the left temporal cortex*, Brain **121** (1998), 1133–1142.
[89] J. Hintikka, *The game of language. Studies in game-theoretical semantics and its applications*, Reidel, 1983.
[90] J. Hintikka, *The emperor's new intuitions*, Journal of Philosophy **96** (1999), 127–147.
[91] P. Hitzler, S. Hölldobler, and A.K. Seda, *Logic programs and connectionist networks*, Journal of Applied Logic **3** (2004), 245–272.
[92] W. Hodges, *From sentence meanings to full semantics*, Unpublished manuscript, 2006.
[93] P.J. Holcomb, *Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing*, Psychophysiology **30** (1993), 47–61.
[94] J. R. Hurford, *The neural basis of predicate-argument structure*, Behavioral and Brain Sciences **26** (2003), 261–316.
[95] J.R. Hurford, *The origins of meaning*, Oxford University Press, 2007.
[96] R. Jackendoff, *Semantics and cognition*, MIT Press, 1983.
[97] ———, *On beyond zebra: The relation of linguistic and visual information*, Cognition **26** (1987), 89–114.
[98] ———, *The architecture of the language faculty*, MIT Press, 1997.
[99] ———, *Foundations of language: Brain, meaning, grammar, evolution*, Oxford University Press, 2002.
[100] ———, *Précis of Foundations of language: Brain, meaning, grammar, evolution*, Behavioral and Brain Sciences **26** (2003), 651–65; discussion 666–707.
[101] ———, *A Parallel Architecture perspective on language processing*, Brain Research **1146** (2007), 2–22.
[102] T. Janssen, *Foundations and Applications of Montague Grammar, Part 1*, Center for Mathematics and Computer Science, Amsterdam, 1986.
[103] ———, *Compositionality*, Handbook of Logic and Language (J. van Benthem and A. ter Meulen, eds.), Elsevier Science, 1997.
[104] K. Johnson, *On the systematicity of language and thought*, Journal of Philosophy **101** (2004), 111–139.
[105] P.N. Johnson-Laird, *Mental models in cognitive science*, Cognitive Science **4** (1980), 71–115.
[106] A. Joshi and Y. Schabes, *Tree-adjoining grammars*, Handbook of formal languages and automata (A. Salomma and G. Rosenberg, eds.), Springer-Verlag, 1997.
[107] H. Kamp, *A theory of truth and semantic representation*, Formal Methods in the Study of Language (J. Groenendijk, T. Janssen, and M. Stokhof, eds.), Amsterdam Mathematical Centre, 1981.
[108] J. Katz and J. Fodor, *The structure of a semantic theory*, Language **39** (1963), 170–210.
[109] A. Kazmi and F.J. Pelletier, *Is compositionality formally vacuous?*, Linguistics and Philosophy **21** (1998), 629–633.
[110] E.L. Keenan, *On surface form and logical form*, Studies in the Linguistic Sciences **8** (1979), 163–203.
[111] A Kim and L Osterhout, *The independence of combinatory semantic processing: Evidence from event-related potentials*, Journal of Memory and Language **52** (2005), 205–225.
[112] J.W. King and M. Kutas, *Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading*, Journal of Cognitive Neuroscience **7** (1995), 376–395.
[113] W. Kohler, *The mentality of apes*, Harcourt Brace and World, 1925.
[114] S. Kripke, *Semantical considerations in modal logic*, Acta Philosophica Fennica **16** (1963), 83–94.
[115] M. Kutas and K.D. Federmeier, *Electrophysiology reveals semantic memory use in language comprehension*, Trends in Cognitive Sciences **4** (2000), 463–470.
[116] M Kutas and S A Hillyard, *Event-related brain potentials to grammatical errors and semantic anomalies*, Memory and Cognition **11** (1983), 539–550.
[117] M. Kutas and S.A. Hillyard, *Reading senseless sentences: Brain potentials reflect semantic incongruity*, Science **207** (1980), 203–205.
[118] ———, *Brain potentials during reading reflect word expectancy and semantic association*, Nature **307** (1984), 161–163.
[119] W. Labov, *When intuitions fail*, Papers from the parasession on theory and data in linguistics. Chicago Linguistic Society **32** (1996), 77–106.
[120] F. Landman, *The progressive*, Natural Language Semantics **1** (1992), 1–32.
[121] R. Larson and G. Segal, *Knowledge of meaning*, MIT Press, 1995.
[122] A. Lascarides and N. Asher, *Temporal interpretation, discourse relations and commonsense entailment*, Linguistics and Philosophy **16** (1993), 437–493.
[123] W. Levelt, *Some psychological aspects of linguistic data*, Linguistische Berichte **17** (1972), 18–30.
[124] S. Levinson, *Presumptive meanings: The theory of generalized conversational implicature*, MIT Press, 2000.
[125] D. Lewis, *General semantics*, Synthese **22** (1970), 18–67.
[126] D.A. Lewis, D.S. Melchitzky, and G.G. Burgos, *Specificity in the functional architecture of primate prefrontal cortex*, Journal of Neurocytology **31** (2002), 265–276.
[127] Xiaoqing Li, Peter Hagoort, and Yufang Yang, *Event-related potential evidence on the influence of accentuation in spoken discourse comprehension in Chinese*, Journal of Cognitive Neuroscience **20** (2008), 906–915.
[128] A. Marantz, *Generative linguistics within the cognitive neuroscience of language*, The Linguistic Review **22** (2005), 429–445.

[129] E. Maris, *Randomization tests for ERP topographies and whole spatiotemporal data matrices*, Psychophysiology **41** (2004), 142–151.

[130] E. Maris and R. Oostenveld, *Nonparametric statistical testing of EEG and MEG data*, Journal of Neuroscience Methods **164** (2007), 177–190.

[131] D. Marr, *Vision: A computational investigation into the human representation and processing of visual information*, Freeman, 1982.

[132] D. Marr and T. Poggio, *From understanding computation to understanding neural circuitry*, MIT AI Memos **357** (1976), 1–22.

[133] W. Marslen-Wilson, *Access and integration: Projecting sound onto meaning*, Lexical representation and process (W. Marslen-Wilson, ed.), MIT Press, 1989, pp. 3–24.

[134] J.J. McCarthy and P.J. Hayes, *Some philosophical problems from the standpoint of artificial intelligence*, Machine Intelligence **4** (1969), 463–502.

[135] J.J. McCarthy and A. Prince, *The emergence of the unmarked: Optimality in prosodic morphology*, Proceedings of NELS, vol. 24, 1994, pp. 333–379.

[136] B. McElree, L. Pylkkanen, M.J. Pickering, and M.J. Traxler, *A time course analysis of enriched composition*, Psychonomic Bulletin and Review **13** (2006), 53–59.

[137] B. McElree, M.J. Traxler, M.J. Pickering, R.E. Seely, and R. Jackendoff, *Reading time evidence for enriched composition*, Cognition **78** (2001), 17–25.

[138] B. McGonigle, M. Chalmers, and A. Dickinson, *Concurrent disjoint and reciprocal classification by* cebus apella *in serial ordering tasks: Evidence for hierarchical organization*, Animal Cognition **6** (2003), 185–197.

[139] R. McKinnon and L. Osterhout, *Constraints on movement phenomena in sentence processing: Evidence from event-related brain potentials*, Language and Cognitive Processes **11** (1996), 495–523.

[140] C.T. McMillan, R. Clark, P. Moore, C. Devita, and M. Grossman, *Neural basis of generalized quantifier comprehension*, Neuropsychologia **43** (2005), 1729–1737.

[141] C.T. McMillan, R. Clark, P. Moore, and M. Grossman, *Quantifier comprehension in corticobasal degeneration*, Brain and Cognition **62** (2006), 250–260.

[142] A. Mecklinger, H. Schriefers, K. Steinhauer, and A.D. Friederici, *Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials*, Memory and Cognition **23** (1995), 477–494.

[143] L. Michaelis, *A unification-based model of aspectual type-shifting*, University of Colorado at Boulder, 2001.

[144] G.A. Miller and P.N. Johnson-Laird, *Language and perception*, Harvard University Press, 1976.

[145] M. Minsky, *Frame-system theory*, Thinking (P.N. Johnson-Laird and P.C. Watson, eds.), Cambridge University Press, 1977, pp. 104–116.

[146] H.M. Müller, J.W. King, and M. Kutas, *Event-related potentials elicited by spoken relative clauses*, Cognitive Brain Research **5** (1997), 193–203.

[147] T.F. Münte, K. Schiltz, and M. Kutas, *When temporal terms belie conceptual order*, Nature **395** (1998), 71–3.

[148] M.S. Nieuwland and J. van Berkum, *When peanuts fall in love: N400 evidence for the power of discourse*, Journal of Cognitive Neuroscience **18** (2005), 1098–1111.

[149] L. Osterhout, *On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences*, Brain and Language **59** (1997), 494–522.

[150] L Osterhout, M Bersick, and J McLaughlin, *Brain potentials reflect violations of gender stereotypes*, Memory and Cognition **25** (1997), 273–285.

[151] L. Osterhout and P. Holcomb, *Event-related brain potentials elicited by syntactic anomaly*, Journal of Memory and Language **31** (1992), 785–806.

[152] ______, *Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech*, Language and Cognitive Processes **8** (1993), 413–437.

[153] L. Osterhout and L.A. Mobley, *Event-related brain potentials elicited by failure to agree*, Journal of Memory and Language **34** (1995), 739–773.

[154] L. Osterhout and J. Nicol, *On the distinctiveness, independence, and time course of brain responses to syntactic and semantic anomalies*, Language and Cognitive Processes **14** (1999), 283–317.

[155] A. Özyürek, R.M. Willems, S. Kita, and P. Hagoort, *On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials*, Journal of Cognitive Neuroscience **19** (2007), 605–616.

[156] A. Papafragou, J. Hulbert, and J. Trueswell, *Does language guide event perception? Evidence from eye movements*, Cognition **108** (2008), 155–184.

[157] T. Parsons, *Events in the semantics of English*, MIT Press, 1990.

[158] B.H. Partee, *Some structural analogies between tenses and pronouns in English*, The Journal of Philosophy **70** (1973), 601–609.

[159] B.H. Partee, A. ter Meulen, and R.E. Wall, *Mathematical methods in linguistics*, Kluwer, 1990.

[160] J. Peregrin (ed.), *Meaning: The dynamic turn*, Elsevier, 2003.

[161] C. Phillips, N. Kazanina, and S.H. Abada, *ERP effects of the processing of syntactic long-distance dependencies*, Cognitive Brain Research **22** (2005), 407–428.

[162] M.M. Piñango, A. Winnick, R. Ullah, and E. Zurif, *The time course of semantic composition: The case of aspectual coercion*, Journal of Psycholinguistic Research **35** (2006), 233–244.

[163] M.M. Piñango, E. Zurif, and R. Jackendoff, *Real-time processing implications of enriched composition at the syntax-semantics interface*, Journal of Psycholinguistic Research **28** (1999), 395–414.

[164] S. Pinker and R. Jackendoff, *The faculty of language: What's special about it?*, Cognition **95** (2005), 201–236.

[165] D. Poeppel and D. Embick, *Defining the relation between linguistics and neuroscience*, Twenty-first century psycholinguistics: Four cornerstones (A. Cutler, ed.), Lawrence Erlbaum Associates, 2005, pp. 103–118.

[166] G. Pullum and B.C. Scholz, *Systematicity and natural language syntax*, Croatian Journal of Philosophy **7** (2007), 375–402.

[167] F. Pulvermuller, *Brain mechanisms linking language and action*, Nature Reviews Neuroscience **6** (2005), 576–582.

[168] F. Pulvermuller, M. Harle, and F. Hummel, *Walking or talking? Behavioral and neurophysiological correlates of action verb processing*, Brain and Language **78** (2001), 143–168.

[169] F. Pulvermuller, O. Hauk, V.V. Nikulin, and R.J. Ilmoniemi, *Functional links between motor and language systems*, European Journal of Neuroscience **21** (2005), 793–797.
[170] K.L. Purvis and R. Tannock, *Language abilities in children with attention deficit disorder, reading disabilities, and normal controls*, Journal of Abnormal Child Psychology **25** (1997), 133–144.
[171] J. Pustejovsky, *The generative lexicon*, MIT Press, 1995.
[172] L. Pylkkänen and B. McElree, *The syntax-semantics interface: On-line composition of sentence meaning*, Handbook of Psycholinguistics (M. Traxler and M.A. Gernsbacher, eds.), Elsevier, 2006.
[173] ______, *An MEG study of silent meaning*, Journal of Cognitive Neuroscience **19** (2007), 1905–1921.
[174] W. V. Quine, *Methodological reflections on current linguistic theory*, Synthese **21** (1970), 386–398.
[175] H. Reichenbach, *Elements of symbolic logic*, Macmillan, 1947.
[176] M. Riesenhuber and T. Poggio, *Are cortical models really bound by the 'binding problem'?*, Neuron **24** (1999), 87–93.
[177] P. Robbins, *Minimalism and modularity*, Context-sensitivity and semantic minimalism (G. Preyer and G. Peter, eds.), Oxford University Press, 2007.
[178] R. Rojas, *Neural Networks: A Systematic Introduction*, Springer, 1996.
[179] F. Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Spartan Books, 1962.
[180] S. Rothstein, *Structuring events*, Blackwell, 2004.
[181] D.S. Ruchkin, R. Jr. Johnson, H.L. Cacoune, W. Ritter, and M. Hammer, *Multiple sources of P3b associated with different types of information*, Psychophysiology **27** (1990), 157–175.
[182] M.D. Rugg, *Dissociation of semantic priming, word and non-word repetition effects by event-related potentials*, Quarterly Journal of Experimental Psychology A **39** (1983), 123–148.
[183] A Sanford, *Context, attention and depth of processing during interpretation*, Mind and Language **17** (2002), 188–206.
[184] A.J. Sanford and P. Sturt, *Depth of processing in language comprehension: not noticing the evidence*, Trends in Cognitive Sciences **6** (2002), 382–386.
[185] R. Schank and R. Abelson, *Scripts, plans, goals and understanding*, Lawrence Erlbaum Associates, 1977.
[186] P.G. Simos, L.F. Basile, and A.C. Papanicolaou, *Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging*, Brain Research **762** (1997), 29–39.
[187] M. Singer, *Discourse inference processes*, Handbook of Psycholinguistics (M.A. Gernsbacher, ed.), Academic Press, 1994.
[188] P. Smolensky and G. Legendre, *The harmonic mind: From neural computation to optimality-theoretic grammar*, MIT Press, 2006.
[189] M.J. Spivey and M. Gonzalez-Marquez, *Rescuing generative linguistics: Too little, too late?*, Behavioral and Brain Sciences **26** (2003), 690–691.
[190] M. Steedman, *Temporality*, Handbook of Logic and Language (J. van Benthem and A. ter Meulen, eds.), Elsevier, 1997, pp. 895–938.
[191] ______, *Plans, affordances, and combinatory grammar*, Linguistics and Philosophy **25** (2002), 723–753.
[192] K. Steinhauer and M. Ullman, *Consecutive ERP effects of morpho-phonology and morpho-syntax*, Brain and Language **83** (2002), 62–65.
[193] K. Stenning, *How did we get here? A question about human cognitive evolution*, Frijda Lecture, University of Amsterdam, 2003.
[194] K. Stenning and M. van Lambalgen, *Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task*, Cognitive Science **29** (2005), 919–960.
[195] ______, *Human Reasoning and Cognitive science*, MIT Press, 2008.
[196] K. Stenning and M. van Lambalgen, *The logical response to a noisy world*, The Psychology of Conditionals (M. Oaksford, ed.), Oxford University Press, 2008.
[197] P. Sturt, *Semantic re-interpretation and garden path recovery*, Cognition **105** (2007), 477–488.
[198] Z.G. Szabó, *Compositionality*, Stanford Encyclopedia of Philosophy, 2007.
[199] L.J. Taylor, S. Lev-Ari, and R.A. Zwaan, *Inferences about action engage action systems*, Brain and Language **107** (2008), 62–67.
[200] I. Toni, F.P. de Lange, M.L. Noordzij, and P. Hagoort, *Language beyond action*, Journal of Physiology-Paris **102** (2008), 71–79.
[201] D.J. Townsend and T.G. Bever, *Sentence comprehension. The integration of habits and rules*, MIT Press, 2001.
[202] T Trabasso and N.L. Stein, *Using goal-plan knowledge to merge the past with the present and future in narrating events on line*, The development of future-oriented processes (M.H. Haith, J.B. Benson, R.J. Roberts, and B.F. Pennington, eds.), University of Chicago Press, 1994, pp. 323–352.
[203] T. Trappenberg, *Fundamentals of computational neuroscience*, Oxford University Press, 2002.
[204] M. Traxler, B. McElree, R. Williams, and M.J. Pickering, *Context effects in coercion: Evidence from eye movements*, Journal of Memory and Language **53** (2005), 1–25.
[205] M.J. Traxler, M.J. Pickering, and B. McElree, *Coercion in sentence processing: Evidence from eye-movements and self-paced reading*, Journal of Memory and Language **47** (2002), 530–547.
[206] J.J. van Berkum, C. Brown, P. Hagoort, and P. Zwitserlood, *Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension*, Psychophysiology **40** (2003), 235–248.
[207] J.J. van Berkum, C.M. Brown, and P. Hagoort, *Early referential contex effects in sentence processing: Evidence from event-related brain potentials*, Journal of Memory and Language **41** (1999), 147–182.
[208] J.J. van Berkum, P. Hagoort, and C.M. Brown, *Semantic integration in sentences and discourse: Evidence from the N400*, Journal of Cognitive Neuroscience **11** (1999), 657–671.
[209] J.J. van Berkum, P. Zwitserlood, P. Hagoort, and C.M. Brown, *When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect*, Cognitive Brain Research **17** (2003), 701–718.
[210] J.J.A. Van Berkum, *Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things?*, The on-line study of sentence comprehension: Eye tracking, ERPs and beyond (M. Carreiras and C. Clifton, eds.), Psychology Press, 2004, pp. 229–270.
[211] T.A. van Dijk, *Macrostructures*, Lawrence Erlbaum Associates, 1980.
[212] M. van Lambalgen and F. Hamm, *The proper treatment of events*, Blackwell, 2004.

[213] M. van Lambalgen and H. Smid, *Reasoning patterns in autism: Rules and exceptions*, Proceedings of the Eighth International Colloquium on Cognitive Science Donostia-San Sebastian (L.A. Perez Miranda and J.M. Larrazabal, eds.), Kluwer, 2004.

[214] M. van Lambalgen, C. van Kruistum, and E. Parriger, *Discourse processing in attention-deficit hyperactivity disorder (ADHD)*, Journal of Logic, Language, and Information **17** (2008), 467–487.

[215] H. van Oostendorp and S. de Mul, *Moses beats Adam: A semantic relatedness effect on a semantic illusion*, Acta Psychologica **74** (1990), 35–46.

[216] C. Van Petten and M. Kutas, *Interactions between sentence context and word frequency in event-related brain potentials*, Memory and Cognition **18** (1990), 380–393.

[217] ———, *Influences of semantic and syntactic context on open- and closed-class words*, Memory and Cognition **19** (1991), 95–112.

[218] Z. Vendler, *Verbs and times*, The Philosophical Review **66** (1957), 143–160.

[219] C. von der Malsburg, *The correlation theory of brain function*, Internal Report 81-82, Department of Neurobiology, Max Planck Institute for Biophysical Chemistry, 1981, Reprinted in: E. Domany, J.L. van Hemmen and K. Schulten (eds.) *Models of neural networks II*, Springer Verlag, 1994.

[220] ———, *The what and why of binding: The modeler's perspective*, Neuron **24** (1999), 95–104.

[221] T. Vosse and G. Kempen, *Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar*, Cognition **75** (2000), 105–143.

[222] D. Westerståhl, *On the mathematical proofs of the vacuity of compositionality*, Linguistics and Philosophy **21** (1998), 635–643.

[223] V.H. Yngve, *A model and a hypothesis for language structure*, Proceedings of the American Philosophical Society **104** (1960), 444–466.

[224] W. Zadrozny, *From compositional to systematic semantics*, Linguistics and Philosophy **17** (1994), 329–342.

Samenvatting

Het is duidelijk dat de hersenen ons in staat stellen te communiceren. Wat minder duidelijk is, is hoe de hersenen precies taal verwerken. Dit boek gaat over de opbouw van complexe betekenissen in het intacte menselijke brein. Om dit te onderzoeken, hebben we modellen uit de formele logica gecombineerd met experimentele methoden uit de neurofysiologie. Daarbij was de vraag hoe mensen tijd, aspect en gebeurtenisstructuur in taal begrijpen het uitgangspunt.

In het eerste deel van het boek, 'From semantics to neuroscience', beschrijven we hoe de semantiek – in het bijzonder de principes van compositionaliteit en monotoniciteit – kan worden gecombineerd met fysiologische metingen zoals ERPs (event-related potentials). In hoofdstuk 1, komt ons onderzoeksplan hiervoor aan de orde. Tevens biedt hoofdstuk 1 een overzicht van de meeste empirische vragen die in hoofdstukken 2-6 worden besproken. Dit hoofdstuk is filosofisch en methodologisch van aard. Het bespreekt de redenen waarom de verschillende disciplines die betekenis onderzoeken, tot nu toe van elkaar gescheiden waren. We suggereren dat de formele semantiek ons in staat stelt het aantal voorspellingen over taalverwerking te beperken. Voorbeelden van deze methodologie worden gegeven in hoofdstukken 1, 2 en 3. Tegelijkertijd kunnen experimentele resultaten worden gebruikt om semantische theorieën te verbeteren.

In het tweede deel van het boek, 'The electrophysiology of meaning', beschrijven we drie ERP-experimenten naar het begrijpen van taal. Deze experimenten waren opgezet om de voorspellingen op basis van de formele analyses uit het eerste deel van het proefschrift te testen. Het eerste experiment (hoofdstuk 4) laat zien dat schendingen van tijdsvorm een negatief ERP effect veroorzaken tussen de 200 en 400 milliseconden na het verschijnen van het werkwoord. Dit effect wijst erop dat de opbouw van betekenissen sterk incrementeel is. In het tweede experiment (hoofdstuk 5), onderzochten we werkwoordelijk aspect. Hier we laten zien dat ERP-resultaten verenigbaar zijn met de niet-monotonische theorie van de 'imperfective paradox' die in hoofdstuk 2 wordt beschreven. We vonden een negatief ERP effect tussen de 400 en 800 milliseconden na het kritieke woord, die was het sterkst over de elektroden in het frontale gebied. We verklaren dit ERP effect in termen van herberekening van betekenis. Ten slotte, in het derde experiment (hoofdstuk 6) hebben we 'complement coercion' onderzocht. We vonden hier bewijs dat er sprake is van een verrijkte semantische samenstelling in de verwerking van gebeurtenisstructuren. Hier tonen de ERP-resultaten een negatief effect die enkele honderden milliseconden duurt. In alle drie de experimenten waren de ERP-effecten van semantische berekening belangrijk verschillend van een N400, die vaak verbonden is aan de verwerking van betekenis in de hersenen.

Dit proefschrift laat zien dat een combinatie van formele modellen en experimentele technieken in het onderzoek naar betekenis van taal waardevol kan zijn. De ERP-resultaten goed verenigbaar zijn met de theorie van taalverwerking die in het eerste deel is beschreven: de opbouw van betekenissen in het brein verloopt incrementeel, en is in wisselwerking met andere niveaus van taalvoorstelling, niet-compositioneel en niet-monotonisch.

Curriculum vitae

Giosuè Baggio was born on April 6, 1979 in Palmanova, Italy. He received his secondary education at Liceo Scientifico Statale 'Giovanni Marinelli' in Udine. He earned a master's degree in philosophy from the University of Pavia in 2002, and a master's degree in logic from the Institute for Logic, Language and Computation of the University of Amsterdam in 2004. In late 2003, he joined the 'Neurocognition of Language' group at the Donders Centre for Cognitive Neuroimaging of the Radboud University Nijmegen, where he worked until early 2009. His research during that period has resulted in the papers listed below, most of which are collected in modified form in this dissertation.

Publications

G. Baggio, T. Choma, M. van Lambalgen & P. Hagoort. Coercion and compositionality. Under review.

G. Baggio, M. van Lambalgen & P. Hagoort. Language, linguistics and cognition. In R. Kempson & T. Fernando (Eds.), *Handbook of Philosophy of Linguistics*. Volume 14 of D. Gabbay, P. Thagard & J.H. Woods (Eds.), *Handbook of Philosophy of Science*. Elsevier, to appear.

P. Hagoort, G. Baggio & R. Willems. Semantic unification. In M.S. Gazzaniga (Ed.), *The New Cognitive Neurosciences*. MIT Press, to appear.

G. Baggio, M. van Lambalgen & P. Hagoort. The processing consequences of compositionality. In W. Hinzen, E. Machery & M. Werning (Eds.), *The Oxford Handbook of Compositionality*. Oxford University Press, to appear.

G. Baggio. Processing temporal constraints: An ERP study. In M. Gullberg & P. Indefrey (Eds.), *Time to Speak: Cognitive and Neural Prerequisites for Time in Language*. Wiley-Blackwell, 2008. Appeared also in *Language Learning* 58, 2008: 35-55.

G. Baggio, M. van Lambalgen & P. Hagoort. Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language* 59, 2008: 36-53.

G. Baggio & M. van Lambalgen. The processing consequences of the imperfective paradox. *Journal of Semantics* 24, 2007: 307-330.

Series Donders Institute for Brain, Cognition and Behaviour

1. van Aalderen-Smeets, S.I. (2007). *Neural dynamics of visual selection*.
Maastricht University, The Netherlands.

2. Schoffelen, J.M. (2007). *Neuronal communication through coherence in the human motor system*.
Radboud University Nijmegen, The Netherlands.

3. de Lange, F.P. (2008). *Neural mechanisms of motor imagery*.
Radboud University Nijmegen, The Netherlands.

4. Grol, M.J. (2008). *Parieto-frontal circuitry in visuomotor control*.
Utrecht University, The Netherlands.

5. Bauer, M. (2008). *Functional roles of rhythmic neuronal activity in the human visual and somatosensory system*.
Radboud University Nijmegen, The Netherlands.

6. Mazaheri, A. (2008). *The influence of ongoing oscillatory brain activity on evoked responses and behaviour*.
Radboud University Nijmegen, The Netherlands.

7. Hooijmans, C.R. (2008). *Impact of nutritional lipids and vascular factors in Alzheimer's Disease*.
Radboud University Nijmegen, The Netherlands.

8. Gaszner, B. (2008). *Plastic responses to stress by the rodent urocortinergic Edinger-Westphal nucleus*.
Radboud University Nijmegen, The Netherlands.

9. Willems, R.M. (2009). *Neural reflections of meaning in gesture, language and action*.
Radboud University Nijmegen, The Netherlands.

10. Van Pelt, S. (2009). *Dynamic neural representations of human visuomotor space*.
Radboud University Nijmegen, The Netherlands.

11. Lommertzen, J. (2009). *Visuomotor coupling at different levels of complexity*.
Radboud University Nijmegen, The Netherlands.

12. Poljac, E. (2009). *Dynamics of cognitive control in task switching: Looking beyond the switch cost*.
Radboud University Nijmegen, The Netherlands.

13. Poser, B.A. (2009). *Techniques for BOLD and blood volume weighted fMRI.*
Radboud University Nijmegen, The Netherlands.

14. Baggio, G. (2009). *Semantics and the electrophysiology of meaning. Tense, aspect, event structure*.
Radboud University Nijmegen, The Netherlands.